NeuroImage 232 (2021) 117917

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Test-retest reliability of emotion regulation networks using fMRI at ultra-high magnetic field

Stella Berboth[a,b,c], Christian Windischberger[c], Nils Kohn[d], Carmen Morawetz[c,e,*]

[a] Department of Neurology, Charité Universitätsmedizin Berlin, Germany
[b] Department of Education and Psychology, Freie Universität Berlin, Germany
[c] Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, Austria
[d] Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Center, Nijmegen, Netherlands
[e] Institute of Psychology, University of Innsbruck, Austria

ARTICLE INFO

Keywords:
Intraclass correlation coefficient
Neuroimaging
Reappraisal
Prefrontal cortex
Emotion generation

ABSTRACT

Given the importance of emotion regulation in affective disorders, emotion regulation is at the focus of attempts to identify brain biomarkers of disease risk, treatment response, and brain development. However, to be useful as an indicator for individual characteristics of brain functions – particularly as a biomarker in a clinical context – ensuring reliability is a key challenge. Here, we systematically evaluated test-retest reliability of task-based functional magnetic resonance imaging (fMRI) activity within neural networks associated with emotion generation and regulation across three sessions. Acquiring fMRI data at ultra-high field (7T), we examined region- and voxel-wise test-retest reliability of brain activity in response to a well-established emotion regulation task for predefined region-of-interests (ROIs) implicated in four neural networks. Test-retest reliability varied considerably across the emotion regulation networks and respective ROIs. However, core emotion regulation regions, including the ventrolateral and dorsolateral prefrontal cortex (vlPFC and dlPFC) as well as the middle temporal gyrus (MTG) showed high reliability. Our findings thus support the role of these prefrontal and temporal regions as promising candidates for the study of individual differences in emotion regulation as well as for neurobiological biomarkers in clinical neuroscience research.

1. Introduction

The ability to regulate our emotions by means of flexibly responding to affective events in a context-dependent manner is of great importance for our mental and physical health (Berking and Wupperman, 2012; Eftekhari et al., 2009) as well as for successful social interactions (Gross and John, 2003). In contrast, impairments in emotion regulation are related to various psychological disorders, such as depression and anxiety (Kring and Sloan, 2010; Sloan et al., 2017). Within the last decade, the role of emotion regulation in the development and maintenance of psychopathologies has been addressed by a growing number of studies, highlighting the importance of emotion dysregulation across psychological disorders (Cludius et al., 2020). Along with the view of emotion regulation as a transdiagnostic construct, understanding the neural mechanisms that underlie the regulation of emotions has become an important topic in affective neuroscience. This is further emphasized in light of the growing interest in finding neurobiological biomarkers of disease development and treatment response (Insel et al., 2010; Woo et al., 2017).

A wealth of neuroimaging studies has investigated the neural substrates of emotion regulation using functional magnetic resonance imaging (fMRI) (Morawetz et al., 2017; Morawetz et al., 2020). Based on several meta-analyses in the field, a well-established and robust network of brain regions has been associated with emotion regulation (e.g., Buhle et al., 2014; Kohn et al., 2014; Morawetz et al., 2020; Ochsner et al., 2012) that consists of frontal (dorsolateral and ventrolateral prefrontal cortex, dlPFC and vlPFC), temporal (superior and middle temporal gyrus, STG and MTG), parietal (inferior and superior parietal lobe) as well as subcortical regions (amygdala, striatum) and the insula.

Notably, the field has evolved from the mere investigation of the neural bases of emotion regulation across healthy and clinical populations to exploring individual differences in emotion regulation and their neural correlates. Research has targeted individual differences in emotion regulation in relation to age (e.g., Silvers et al., 2017), reappraisal success (e.g., Morawetz et al., 2016), personality traits (e.g., S. Chen et al., 2017; Morawetz et al., 2017), physiological measures (e.g., Urry et al., 2006) or psychiatric symptoms (e.g., Reinecke et al., 2015). For classical fMRI research, many paradigms were optimized by mini-

* Corresponding author at: Institute of Psychology, University of Innsbruck, Innrain 52f, 6020 Innsbruck, Austria.
E-mail address: carmen.morawetz@uibk.ac.at (C. Morawetz).

https://doi.org/10.1016/j.neuroimage.2021.117917
Received 2 September 2020; Received in revised form 29 January 2021; Accepted 22 February 2021
Available online 27 February 2021

mizing between-subject variability to obtain large effects at the group level resulting in robust brain responses across studies (Elliott et al., 2020; Fröhner et al., 2019). In contrast, the study of individual differences relies on the reproducible assessment of differences between subjects. With this shift of research focus new statistical challenges arise (Dubois and Adolphs, 2016): To be useful as an indicator for individual characteristics of brain functions – particularly as a biomarker in a clinical context – one key challenge is to ensure the reliability of the measure. Reliability reflects the ability of a measure to give the same results under similar circumstances, i.e. the ability to measure consistent results across repeated measurements in terms of test-retest reliability. It represents a prerequisite for the validity, power and predictive utility of a measure (Elliott et al., 2020).

In light of the ongoing replicability crisis ("Fostering reproducible fMRI research," 2017), the discussion on task fMRI reliability has received new momentum, resulting in a surge of interest in fMRI reliability studies (Bennett and Miller, 2010; Elliott et al., 2020; Noble et al., 2019; Vul et al., 2009). However, the overall reliability of commonly used fMRI tasks remains largely unknown as reliability estimates of individual studies vary widely between tasks as well as between studies investigating the same fMRI task (Elliott et al., 2020). Hence, it is essential to assess and report reliability for each specific task and scenario. The evaluation of the reliability of emotion regulation paradigms, that are widely used to assess underlying neural processes on a group and individual level, is therefore a key challenge within affective neuroscience, that has not been addressed so far.

In the current study, we therefore aimed to (1) determine the reliability and thus the applicability of a well-established emotion regulation task to assess individual differences and (2) identify regions within emotion regulation networks that can be reliably measured, therefore qualifying as promising biomarker candidates for future studies. To address these issues, we used a well-established emotion regulation task, that represents the most commonly used task within the field (Morawetz et al., 2017). Typically, a negative picture is presented as an emotional stimulus and participants are asked to either down-regulate their emotion or to look at the image and let themselves respond naturally, allowing the differentiation of emotion generative from emotion regulatory processes by contrasting the different conditions (Gross et al., 2011; McRae et al., 2012; Otto et al., 2014). In this study, the emotion regulation task was conducted at three time points separated by one week using ultra high-field fMRI (7T), allowing for the assessment of test-retest reliability for two different time intervals: a shorter, one-week and a longer, two-weeks' time interval. Using this design, the present study tested for the first time the reliability of emotion regulation ability on a behavioral and neural level. To assess local reliability, we focused on neural networks that have been identified in a recent meta-analysis including over 100 neuroimaging studies. In this meta-analysis four dissociable yet interacting large-scale networks have been determined, which have been linked to different psychological processes (Morawetz et al., 2020). The first two networks (N1 and N2) have mainly been associated with emotion regulation and include frontal and parietal cortical regions related to response inhibition or executive control, attention, memory and language. The other two networks, based upon mainly subcortical regions, have been related to the perception of different emotional qualia and seem to play a central role in emotional reactivity and the generation of emotional responses (N3), and the perception and processing of internal sensations (N4). These networks were used as *a priori* defined regions-of-interest (ROIs) to assess test-retest reliability in an independent manner and to evaluate the applicability of the measure on an individual level. Here, we sought to thoroughly investigate the reliability of the respective ROIs by performing test-retest reliability analyses on two levels: Firstly, we assessed region-wise reliability based on the mean ROI activation for each participant. Secondly, we further explored reliability patterns by calculating ICC maps on a whole-brain level and extracting voxel-wise reliability for the respective ROIs.

We hypothesized that emotion regulation would be associated with increased activity within dlPFC, vlPFC, MTG, STG, and the inferior and superior parietal lobe within all three sessions. We further expected that brain activity would habituate over time (e.g. Plichta et al., 2012; Sauder et al., 2014). In relation to the test-retest reliability of the four predefined networks (N1 – N4), implicated in emotion generation and regulation, we hypothesized that test-retest reliability would vary between regions. In particular, we expected higher test-retest reliability in prefrontal regions given the stronger activation patterns across sessions compared to subcortical regions which are more prone to susceptibility artifacts and less activated in general (Elliott et al., 2020). More specifically, we predicted that emotion regulation networks (N1 and N2), mainly consisting of prefrontal and temporal regions, would demonstrate higher reliability than the other two networks (N3 and N4) including subcortical regions such as the amygdala.

As of today, no study has systematically examined the effect of repeated measurements on the assessment of emotion regulation on a behavioral and neural level. Thus, this study represents a first step to determine the reliability of emotion regulation and the underlying task-based fMRI activity within emotion regulation networks across multiple sessions. Using different indexes of reliability, this work not only provides evidence for the reliability of brain responses elicited by an emotion regulation task for the study of individual differences, but also highlights which brain regions might serve as promising brain biomarkers for future studies.

2. Materials and methods

2.1. Participants

32 participants (27 females, mean age = 22.56 years, *SD* = 3.06, range = 19–35) were recruited to participate in the fMRI experiment via advertisement at online platforms such as facebook and mailing lists of the University of Vienna and the Medical University of Vienna. Seven participants had to be excluded because they did not return for the second and third MRI session ($n = 4$), or due to excessive head movement (movement >3 mm/ >3° in one direction; $n = 3$). The final sample consisted of 25 right-handed, healthy participants with normal or corrected to normal vision (21 females, mean age = 22.80 years, *SD* = 3.30, range = 19–35). All participants gave written, informed consent and reported no history of neurological or psychiatric disorders.

The study was approved by the local ethics committee of the Medical University of Vienna and carried out in accordance with the Declaration of Helsinki.

2.2. Experimental design

2.2.1. Procedure

All participants were tested three times. The three fMRI sessions were separated by one week in 76% of the cases. Due to illness or technical problems, few participants exceeded these time intervals: One participant returned 14 days after the first session and 6 participants returned on average within 15 ± 2 days after the second session. Each fMRI session involved four runs of the emotion regulation task outlined below. In addition, each fMRI session was followed by a rating experiment, in which participants were instructed to rate the stimuli of the preceded emotion regulation task on valence and arousal.

2.2.2. Stimuli

Stimuli consisted of 240 aversive images from the International Affective Picture System (IAPS, $n = 81$ pictures; mean arousal = 6.17, *SD* = 0.62; mean valence = 2.26, *SD* = 0.59) (Bradley and Lang, 2007) and the Nencki Affective Picture System (NAPS, $n = 159$ pictures; mean arousal = 6.42, *SD* = 0.63; mean valence = 3.13, *SD* = 0.64) (Marchewka et al., 2014). Images were assigned to the two task conditions (*Look* or *Decrease*, see below) and split into three sessions.

These image sets were matched in content and did not differ in normative arousal and valence ratings to ensure that emotion induction was comparable between the conditions and sessions (session 1: mean arousal = 6.36, *SD* = 0.79, mean valence = 2.88, *SD* = 0.71; session 2: mean arousal = 6.31, *SD* = 0.74, mean valence = 2.84, *SD* = 0.67; session 3: mean arousal = 6.34, *SD* = 0.71, mean valence = 2.79, *SD* = 0.54; normative valence and arousal values are indicated on a nine-point Likert scale from 1 (very positive/calm) to 9 (very negative/highly arousing)). However, we intended to use a wide spectrum of negative images and included high as well as low arousing images evenly distributed across all three sessions.

After each fMRI experiment, the stimuli were rated on valence and arousal on a nine-point Likert scale from 1 (very positive/calm) to 9 (very negative/highly arousing) using the Self Assessment Manikin (Bradley and Lang, 2007). An additional set of eight aversive pictures from the IAPS database was used during the practice session outside the scanner. During the fMRI experiment, images were presented in the center of the screen with an 800 × 600 pixel display using the stimulation software Presentation (Version 20.2, Neurobehavioral Systems, USA).

2.2.3. Emotion regulation task

The task design was adapted from a well-established emotion regulation task (e.g., Denny et al., 2015; Morawetz et al., 2017; Ochsner et al., 2004). Two task conditions were implemented in the experiment. In the *Look* condition, participants were instructed to look at the picture, respond naturally to the presented image, and allow themselves to experience any emotional responses without trying to manipulate them. In contrast, in the *Decrease* condition, participants were asked to reduce the intensity of the negative emotion by using reappraisal as emotion regulation strategy. Participants were instructed to reduce the negative impact of the depicted situation by distancing themselves from the image by becoming a detached observer, taking the perspective of a professional observer, or thinking that the depicted situation is not real. Before the first fMRI measurement, participants received a training session to practice the reappraisal strategy (8 trials).

The paradigm was implemented as an event-related design (**Fig. 1**). During the task, stimuli were displayed using a video projector on a screen at the head end of the scanner. Each trial started with the instruction cue (2 s) showing either the words “Down-regulate” (*Decrease*) or “Maintain” (*Look*) at the center of the screen. After the instruction, a jittered fixation cross was presented (2–6 s) followed by the aversive image (8 s). Subsequently, another jittered fixation cross (2–6 s) was presented, followed by a rating of the current emotional state (4 s) on a continuous scale from −200 to +200. The extremes of the rating scale were labeled ‘very negative’ and ‘very positive’, respectively. Participants indicated their emotional state by using a two-button MR-compatible response pad (Current Designs, Philadelphia, PA) to navigate a cursor along the rating scale. The cursor was presented in the middle of the scale on each trial. Finally, the trial concluded by a jittered fixation cross (2–6 s). One experimental run consisted of 20 trials (10 per task condition). Trials were randomized within each run. Each experimental session consisted of four runs resulting in 80 trials per session and 240 trials in total for all three sessions.

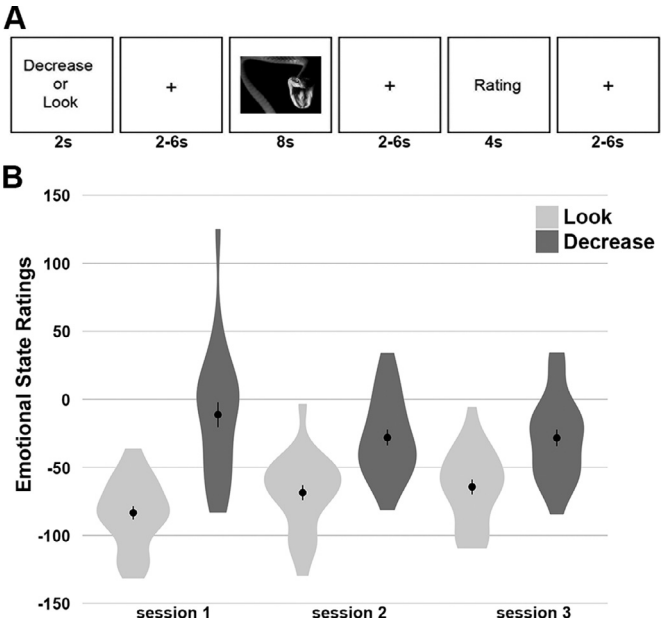

Fig. 1. (A) Task design. Each trial started with an instruction screen of 2 s, showing either the words “Down-regulate” or “Maintain” indicating the experimental condition *Decrease* and *Look*, respectively. The instruction was followed by a jittered fixation cross (2–6 s). Then an aversive picture was presented for 8 s, during which participants were asked to either down-regulate (*Decrease*) their emotions or not modulate their emotions at all (*Look*), followed by a jittered fixation phase (2–6 s). After this regulation phase, participants rated their current emotional state on a continuous scale from very negative to very positive within 4 s. Each trial ended with a jittered fixation phase of 2–6 s. **(B)** Emotional state ratings as a function of task condition and session. Error bars represent standard errors.

2.3. fMRI data acquisition

Whole-brain functional and anatomical images were acquired using a SIEMENS Magnetom 7.0 Tesla MR scanner and a 32-channel head coil. A high-resolution 3D T1-weighted dataset was acquired for each participant and each session (224 sagittal sections, 0.7 × 0.7 × 0.7 mm^3; 224 × 224 data acquisition matrix). Functional images were acquired using the multiband EPI sequence (Version R016a; TR = 1.4 s; TE = 23 ms; 78 slices; voxel size = 1.5 × 1.5 × 1.0 mm^3; 0.2 mm slice gap; field of view = 192 × 192 × 97.5 mm^3; flip angle = 62°) developed at the Center for Magnetic Resonance Research (CMRR, University of Minnesota, Minnesota, USA) (Moeller et al., 2010). For each experimental run, 371 whole-brain volumes were recorded.

Note that despite the advantages of ultra-high magnetic field MRI, including increased image and time course signal-to-noise ratio, spatial resolution, and higher BOLD-related signal changes (Balchandani and Naidich, 2015; Moser et al., 2012), higher field strength also results in increased signal dropouts along susceptibility borders that originate from heterogeneous fields within a voxel (Moser et al., 2012; Triantafyllou et al., 2005). Ventral brain areas, including subcortical structures as well as the orbitofrontal and temporal cortex, are particularly affected by susceptibility-related effects (Merboldt et al., 2001). To compensate for these challenges, we applied a high-resolution scanning protocol, including advanced shimming procedures, small voxel sizes and parallel imaging, leading to reduced intravoxel field inhomogeneities and thus reduced dephasing and signal dropout (Balchandani and Naidich, 2015; Geissberger et al., 2020; Robinson et al., 2004; Sladky et al., 2013; Triantafyllou et al., 2005).

2.4. Data analyses

2.4.1. Reliability assessment of behavioral performances and brain activation

To analyze the test-retest reliability of behavioral and fMRI data we calculated intraclass correlation coefficients (ICCs). The ICC represents a standard method to quantify test-retest reliability (Bennett and Miller, 2010; Noble et al., 2020) and is typically interpreted as the ratio of variance of interest and total variance (Shrout and Fleiss, 1979). The ICC values range from 1.0 indicating near-perfect agreement between the values of the test and retest measure to 0 indicating no agreement between the values of test and retest measure.

We assessed test-retest reliability by calculating consistency ICC estimates based on a single-measurement, two-way mixed effects model (McGraw and Wong, 1996) referring to ICC(3,1) as defined by Shrout and Fleiss (1979). The total sum of squares of this two-way ANOVA is split into a between-subject sum of squares, a between-session sum of squares and a residual sum of squares (corresponding mean squares: between-subjects mean square (BMS), between-sessions mean square (JMS, the original terminology of "J" is "Judge" as used by Shrout and Fleiss (1979)) and error mean square (EMS)). In the mixed effects model, the effect of subjects is assumed to be random while the effect of session is modeled as a fixed effect, as test-retest measurements are considered to show a systematic source of variance (G. Chen et al., 2018; Koo and Li, 2016; McGraw and Wong, 1996). Due to the possibility of habituation or attenuation effects over sessions (e.g., Geissberger et al., 2020; Plichta et al., 2012; Sauder et al., 2014), we employed the consistency type of ICC. Thus, instead of measuring agreement in absolute values, we assessed the extent of agreement of the values of participants after accounting for potential systematic differences such as habituation, attenuation or training effects. In terms of task-based fMRI data, a high ICC indicates that a participant's brain region demonstrating higher activation in session 1 relative to the rest of the sample also shows higher activation than the rest of the sample in session 2. For the consistency type of ICC, only the BMS and EMS are taken into account:

$$ICC(3,1) = \frac{BMS - EMS}{BMS + (k-1)EMS} \quad (1)$$

where k = number of repeated sessions.

We used single-measurement ICCs as the effect estimates were derived from a single session. To quantify the degree of reliability a common guideline for interpretation was applied: ICCs below 0.40 reflect poor reliability, ICCs between 0.40 and 0.59 fair, ICCs between 0.60 and 0.75 good, and ICCs higher than 0.75 excellent reliability (Cicchetti and Sparrow, 1981).

Note, that although ICC values are theoretically defined as nonnegative, they can be negative and therefore uninterpretable, but often are interpreted as poor to zero reliability (Bartko, 1976; Giradeau, 1996; Lahey et al., 1983). Regarding neuroimaging studies, negative ICCs can actually occur for a large number of voxels (G. Chen et al., 2018). Therefore, negative values reported in this study cannot be interpreted.

2.4.2. Analysis of self-reported data

2.4.2.1. Control analyses: stimulus features. Based on the stimulus ratings after each fMRI session, mean arousal and valence ratings were calculated for each participant in each condition and session to ensure that participants perceived the pictures as negative and arousing throughout the experiment.

2.4.2.2. Emotional state ratings and regulation success. Mean emotional state ratings were calculated for the two conditions and each session. In a first step, we performed a repeated-measures ANOVA with the factors condition (*Look, Decrease*) and session (session 1, session 2, session 3) to analyze the effects of task conditions and session on emotional state ratings, followed by post-hoc paired t-tests (two-tailed; p-values Bonferroni-corrected). In a second step, regulation success scores were calculated using the emotional state ratings. For this, the mean difference between the emotional state ratings during *Look* trials and during *Decrease* trials was calculated. These difference scores, reflecting regulation success, are typically calculated and applied for the investigation of individual differences in emotion regulation (e.g., McRae et al., 2012; Otto et al., 2014; Wager et al., 2008).

2.4.2.3. Reliability analyses of behavioral performance. As fMRI provides a method for measuring neural activation in association with cognitive and affective tasks and the corresponding behavioral response, the reliability of behavioral data might influence the reliability of the fMRI measurement. Therefore, we analyzed the reliability of behavioral data by calculating ICCs in R (R package psych; Revelle, 2019). We performed ICC(3,1) analysis on mean emotional state ratings for the *Decrease* and *Look* condition, respectively, as well as for regulation success scores. ICCs and their corresponding confidence intervals (CI) were calculated for every possible combination of time points: session 1 | session 2, session 2 | session 3, session 1 | session 3.

2.4.3. fMRI data analysis

2.4.3.1. Preprocessing. Functional MRI data were analyzed using SPM12 (Statistical Parametric Mapping, Wellcome center for Human Neuroimaging, London, UK) in Matlab R2018b (MathWorks, Natick, MA, USA). Preprocessing included slice-time correction, realignment, coregistration to the respective structural image of the participant, spatial normalization to the standard EPI template (Montreal Neurological Institute, MNI), and smoothing with an isotropic Gaussian kernel (6 mm full-width at half-maximum). Single runs of the emotion regulation task in which participants showed extensive head movement (>3 mm/ >3° in one direction) were excluded from further analyses resulting in the exclusion of twelve runs ($n = 4$ runs in session 1, $n = 4$ runs in session 2, $n = 4$ runs in session 3). Note, that despite the exclusion of single runs, a minimum of 2 runs was available for each participant in each session.

2.4.3.2. Whole-brain group level analyses. The first-level model consisted of stimulus onset vectors representing the instruction (duration 2 s), the emotion regulation phase during stimulus viewing split by task conditions (*Look, Decrease*) (duration 8 s), and the rating phase (4 s) that were convolved with the haemodynamic response function. Six movement parameters were included in the model as nuisance regressors. The regressors-of-interest were the two task conditions *Look* and *Decrease*.

The resulting single-subject contrast images for *Look* and *Decrease* were submitted to the second-level group-analysis. A random-effects general linear model (GLM) was computed using a flexible factorial design. The model consisted of two factors: task condition (*Look, Decrease*) and session (*session 1, session 2, session 3*). Contrasts were computed to test for effects of emotion regulation (*Decrease* > *Look*) and emotion generation (*Look* > *Decrease*) as well as the effect of session (*session1* > *session2* > *session3* and *session3* > *session2* > *session1*). T-statistics were thresholded at an initial cluster-defining threshold $p < .001$, corrected for multiple comparisons with family wise error rate (FWE) at $p < .05$. Coordinates of the results are reported in MNI space.

2.4.3.3. Region-of-interest (ROI) definition. In the present study we used independent, *a priori* defined ROIs from a recent meta-analysis on emotion regulation networks (Morawetz et al., 2020), which represents a commonly applied method for conducting ROI analyses and an appropriate method for ROI analyses in individual differences research (Kriegeskorte et al., 2009). Morawetz et al. (2020) identified four large-scale networks in a meta-analytic grouping approach of fMRI experiments investigating emotion regulation. Two of these networks (N1 and N2) are associated with emotion regulation and include frontal and parietal cortical regions related to response inhibition or executive control, attention, memory and language. The other two networks including mainly subcortical regions are related to the perception of different emotional qualia and seem to play a central role in emotional reactivity and the generation of emotional responses (N3), and the perception and processing of internal sensations (N4). In sum, this resulted in 34 ROIs clustered into the four networks (N1: 10 ROIs, N2: 8 ROIs, N3: 7 ROIs, and N4: 9 ROIs) that were used to assess reliability during the emotion regulation task. Details of the ROIs are reported in the Supplementary Material, **Table S1**.

2.4.3.4. Reliability analysis of fMRI data. We used two approaches to assess the test-retest reliability of the task-based fMRI data, *region-wise* and *voxel-wise* reliability. Both measures are commonly used to investigate the reliability of task-fMRI (e.g. Geissberger et al., 2020; Li et al., 2020; Lois et al., 2018) and are highly, but not perfectly, correlated

(Caceres et al., 2009; Fröhner et al., 2019). To assess *region-wise* reliability, the mean ROI activation for each participant and each ROI was extracted using Marsbar (Toolbox for SPM; Brett et al., 2002). ICC values and their confidence intervals (CI) were then calculated for the mean activation of all ROIs using R. Next, we sought to assess *voxel-wise* reliability, allowing further visualization of the reliability pattern within the networks and ROIs, respectively. For this, we calculated ICC maps on the whole-brain level using the fmreli toolbox implemented in SPM (Fröhner et al., 2019). We then extracted the distribution of voxel-wise ICC values within each ROI and obtained the medians of the ICC distribution for each region, which constitute a measure of test-retest reliability at the voxel level (Caceres et al., 2009), using the statistical program R (Version 3.3.0; R CoreTeam, 2016).

As the reliability of main contrasts and difference contrasts was shown to differ due to shared true score variance of conditions (Infantolino et al., 2018), we here report reliability measures for both, the difference contrast (i.e. *Decrease* > *Look*) that was of main interest, as well as for the contrasts of main effects (i.e. *Decrease* > *Baseline; Look* > *Baseline*).

3. Results

First, we report the behavioral results as well as the test-retest reliability analyses of the emotion regulation task. Second, task effects across all sessions and for each session separately are reported on a whole-brain level. Third, to determine the reliability of the emotion regulation task on a neural level we conducted *region-wise* and *voxel-wise* ICC analyses using the four predefined networks shown to be involved in emotion generation and regulation.

3.1. Behavioral results

3.1.1. Control analysis: stimulus features

Participants were asked to rate all pictures after each session on valence and arousal. Results confirmed that the pictures induced the intended negative affect: For each session, participants rated the stimuli as negative and arousing in the *Decrease* (session 1: mean valence = 3.15, *SD* = 0.57, mean arousal = 4.72, *SD* = 0.93; session 2: mean valence = 3.09, *SD* = 0.71, mean arousal = 5.04, *SD* = 1.03; session 3: mean valence = 3.19, *SD* = 0.63, mean arousal = 4.94, *SD* = 1.01) and in the *Look* condition (session 1: mean valence = 3.13, *SD* = 0.50, mean arousal = 4.81, *SD* = 0.90; session 2: mean valence = 3.17, *SD* = 0.66, mean arousal = 4.89, *SD* = 1.10; session 3: mean valence = 3.25, *SD* = 0.66, mean arousal = 4.70, *SD* = 0.91).

3.1.2. Emotional state ratings

Analyses of the emotional state ratings revealed a significant main effect of emotion regulation condition ($F(1,24) = 51.04$, $p < .001$), no significant main effect of session ($F(2,48) = 0.17$, $p = .84$) and a significant interaction effect between emotion regulation condition and session ($F(1.50,36.06) = 20.87$, $p < .001$, Greenhouse-Geisser corrected) (**Fig. 1**). *Post-hoc* paired t-tests showed that participants felt significantly less negative during *Decrease* compared to the *Look* condition in each session (session 1: $t(24) = 6.79$, $p < .001$, $d = 1.36$; session 2: $t(24) = 6.96$, $p < .001$, $d = 1.39$, session 3: $t(24) = 6.11$, $p < .001$, $d = 1.22$). Further, within the *Look* condition ratings for session 1 differed significantly from session 2 ($t(24) = -3.85$, $p = .01$, $d = 0.77$) as well as from session 3 ($t(24) = -6.95$, $p < .001$, $d = 1.39$). Within the *Decrease* condition, ratings did not differ significantly between sessions. All *post-hoc* pairwise comparisons are reported in **Table 1**.

3.1.3. Reliability of emotional state ratings and regulation success

To assess reliability of behavioral data between the sessions, we performed ICC analyses on the behavioral performance estimates, i.e. the emotional state ratings and the regulation success respectively. For the emotional state ratings we calculated ICC(3,1) values separately for the *Decrease* and *Look* condition that were in the good to excellent range for the *Decrease* condition ($ICC_{s1\text{-}s2} = 0.64$, $p<.001$, CI = [0.34, 0.83]; $ICC_{s2\text{-}s3} = 0.89$, $p<.001$, CI = [0.77, 0.95]; $ICC_{s1\text{-}s3} = 0.66$, $p<.001$, CI = [0.37, 0.84]) and the *Look* condition ($ICC_{s1\text{-}s2} = 0.74$, $p<.001$, CI = [0.49, 0.87]; $ICC_{s2\text{-}s3} = 0.79$, $p<.001$, CI = [0.58, 0.90]; $ICC_{s1\text{-}s3} = 0.86$, $p<.001$, CI = [0.71, 0.94]). ICC analyses of the regulation success scores revealed good to excellent reliability for each combination of session ($ICC_{s1\text{-}s2} = 0.65$, $p<.001$, CI = [0.40, 0.81]; $ICC_{s2\text{-}s3} = 0.76$, $p<.001$, CI = [0.58, 0.87]; $ICC_{s1\text{-}s3} = 0.70$, $p<.001$, CI = [0.49, 0.84). Note, that the regulation success is the result of a subtraction of the emotional state ratings of the *Look* and *Decrease* condition. Thus, in comparison to the emotional state ratings, shared variance in the data gets lost, resulting in slightly lower reliability compared to the emotional state ratings (Hedge et al., 2018).

Table 1
Post-hoc t-tests of emotional state ratings.

Comparison	t (df = 24)	p value	Cohen's d
$Look_{s1}$ > $Decrease_{s1}$	−6.79	**<0.001**	1.36
$Look_{s2}$ > $Decrease_{s2}$	−6.96	**<0.001**	1.39
$Look_{s3}$ > $Decrease_{s3}$	−6.11	**<0.001**	1.22
$Look_{s1}$ > $Look_{s2}$	−3.85	**0.01**	0.77
$Look_{s1}$ > $Look_{s3}$	−6.95	**<0.001**	1.39
$Look_{s2}$ > $Look_{s3}$	−1.21	1	0.24
$Decrease_{s1}$ > $Decrease_{s2}$	2.58	0.25	0.52
$Decrease_{s1}$ > $Decrease_{s3}$	2.65	0.21	0.53
$Decrease_{s2}$ > $Decrease_{s3}$	0.09	1	0.02
$Look_{s1}$ > $Decrease_{s2}$	−8.38	**<0.001**	1.68
$Look_{s1}$ > $Decrease_{s3}$	−8.34	**<0.001**	1.67
$Look_{s2}$ > $Decrease_{s1}$	−5.49	**<0.001**	1.10
$Look_{s2}$ > $Decrease_{s3}$	−6.33	**<0.001**	1.27
$Look_{s3}$ > $Decrease_{s1}$	−5.07	**<0.001**	1.01
$Look_{s3}$ >$Decrease_{s2}$	−6.40	**<0.001**	1.28

Note. Results are Bonferroni corrected for multiple comparisons (n = 15). Significant results are indicated in bold. df=degrees of freedom, s = session.

3.2. fMRI results

3.2.1. Whole-brain analyses

First, we investigated activity changes due to emotion generation and regulation across all sessions on the whole-brain level. Consistent with previous studies (Kohn et al., 2014; Morawetz et al., 2017), testing for the effects of emotion regulation [*Decrease* > *Look*] and emotion generation [*Look* > *Decrease*] across the three sessions revealed increased activity in a widespread network of lateral prefrontal, temporal, and parietal regions. Details of the results are reported in the Supplementary Material (**Table S2** and **Fig. S1**). Testing for session effects resulted in no significant clusters of activation on the whole-brain level, suggesting that activation did not significantly vary over time.

In a second step, we tested for task effects for each session separately. The previously observed regions were robustly activated during emotion generation and regulation across sessions (**Table 2** and **Fig. 2**). However, in general for the emotion regulation contrast [*Decrease* > *Look*] activation declined over time (**Fig. 2**). During the first session, increased activity during the down-regulation was found in frontal (right IFG and SFG, left middle frontal gyrus (MFG), SMA, and precentral gyrus), temporal (bilateral MTG and STG), and parietal regions (bilateral angular gyrus and the precuneus) (**Fig. 2A**). In the second session the frontal (right SFG, left IFG and SMA, bilateral MFG), temporal (left MTG) and parietal cortices (bilateral angular gyrus and inferior parietal lobe (IPL)) demonstrated enhanced responses during emotion regulation (**Fig. 2B**). While in the third session, increased activation was constrained to frontal (left MFG and precentral gyrus) and temporal regions (right STG, bilateral MTG) (**Fig. 2C**).

The emotion generation contrast [*Look* > *Decrease*] revealed enhanced activity in the left STG, bilateral supramarginal gyrus, and bilateral insulae during the first session (**Fig. 2D**). In the second session

Table 2
Brain regions involved in emotion generation and regulation for each session separately (whole-brain analyses).

Region	Side	Coordinates			BA	Cluster size	*t* value	*p* value
		x	y	z				
Session 1: Decrease > Look								
Superior Temporal Gyrus	L	−50	10	−24	21	1258	6.28	<0.001
Middle Temporal Gyrus	L	−56	−1	−22	21		4.63	
Middle Temporal Gyrus	L	−51	−10	−11	22		4.52	
Superior Frontal Gyrus	R	20	16	60	8	6923	6.2	<0.001
Superior Frontal Gyrus	R	20	11	67	6		5.52	
Supplementary Motor Area	L	−4	11	54	6		5.52	
Angular Gyrus	R	54	−54	30	39	4939	6.16	<0.001
Angular Gyrus	R	45	−52	22	39		5.79	
Superior Temporal Gyrus	R	57	−50	22	22		4.83	
Middle Temporal Gyrus	L	−52	−43	−2	21	3525	5.9	<0.001
Angular Gyrus	L	−58	−58	27	22		5.14	
Angular Gyrus	L	−50	−62	45	39		4.69	
Superior Frontal Gyrus	R	20	48	33	9	1026	5.24	<0.001
Superior Frontal Gyrus	R	18	58	30	9		4.2	
Middle Temporal Gyrus	R	45	−37	2	21	763	5.19	0.003
Middle Temporal Gyrus	R	56	−32	−2	21		4.33	
Precuneus		6	−50	48		937	4.98	0.001
Precuneus		−10	−52	43			3.58	
Inferior Frontal Gyrus	R	48	36	−12	47	649	4.71	0.007
Inferior Frontal Gyrus	R	51	26	−8	38		4.64	
Inferior Frontal Gyrus	R	57	23	14	45		4.64	
Precentral Gyrus	L	−34	−2	63	6	1265	4.52	<0.001
Middle Frontal Gyrus	L	−39	4	54	6		4.32	
Precentral Gyrus	L	−38	6	44	6		4.02	
Session 2: Decrease > Look								
Supplementary Motor Area	L	−3	14	60	6	6725	5.53	<0.001
Middle Frontal Gyrus	L	−38	20	42	46		5.2	
Superior Frontal Gyrus	R	18	20	63	8		5.04	
Angular Gyrus	R	52	−56	28	22	1384	4.89	<0.001
Inferior Parietal Gyrus	R	48	−54	43	40		3.86	
Angular Gyrus	L	−46	−66	45	39	1415	4.74	<0.001
Angular Gyrus	L	−44	−58	40	39		4.24	
Inferior Parietal Gyrus	L	−51	−49	55	40		3.88	
Inferior Frontal Gyrus	L	−46	30	−12	38	1400	4.7	<0.001
Middle Frontal Gyrus	L	−36	53	−2	47		4.45	
Inferior Frontal Gyrus	L	−36	41	−14	47		4.44	
Superior Frontal Gyrus	R	15	58	31	9	1895	4.26	<0.001
Middle Frontal Gyrus	R	34	20	48	9		4.16	
Middle Frontal Gyrus	R	39	24	42	46		4.05	
Middle Temporal Gyrus	L	−56	−43	0	21	577	4.11	0.012
Middle Temporal Gyrus	L	−64	−42	2	22		4.09	
Session 3: Decrease > Look								
Middle Temporal Gyrus	R	52	−67	12	37	3707	5.7	<0.001
Middle Temporal Gyrus	R	51	−60	16	37		4.97	
Superrior Temporal Gyrus	R	57	−48	21	22		4.78	
Middle Temporal Gyrus	L	−51	−74	6	19	1847	4.86	<0.001
Middle Temporal Gyrus	L	−46	−66	12	37		4.61	
Middle Temporal Gyrus	L	−56	−50	20	22		4.02	
Precentral Gyrus	L	−40	−2	61	6	456	3.87	0.033
Middle Frontal Gyrus	L	−36	4	55	6		3.83	
Precentral Gyrus	L	−32	−2	58	6		3.78	
Session 1: Look > Decrease								
Insula	R	40	−13	−4	48	5303	6.62	<0.001
Insula	R	40	−1	−11	48		5.81	
Supramarginal Gyrus	R	54	−28	28	48		5.22	
Insula	L	−39	−18	2	48	3604	6.59	<0.001
Insula	L	−40	−6	−5	48		5.92	
Insula	L	−34	2	15	48		5.67	
Superior Temporal Gyrus	L	−64	−30	21	48	2028	5.44	<0.001
Supramarginal Gyrus	L	−60	−26	28	48		5.07	
Supramarginal Gyrus	L	−60	−22	15	42		4.98	
Session 2: Look > Decrease								
Superior Temporal Gyrus	L	−40	−24	3	48	8722	5.86	<0.001
Cingulate Gyrus		−22	−37	32			5.4	

(*continued on next page*)

Table 2 (*continued*)

Region	Side	Coordinates			BA	Cluster size	*t* value	*p* value
		x	y	z				
Angular Gyrus	L	−22	−48	28			5.34	
Insula	R	44	−13	2	48	4797	5.73	<0.001
Rolandic Operculum	R	51	−25	22	48		5.62	
Supramarginal Gyrus	R	58	−26	19	42		5.4	
Precuneus		22	−54	36		1352	4.3	<0.001
Postcentral Gyrus	R	24	−30	78	4		4.16	
Paracentral Lobule	R	14	−31	54			4.15	
Session 3: Look > Decrease								
Insula	L	−22	34	7	47	1476	4.84	<0.001
Inferior Frontal Gyrus	L	−16	38	−6	11		4.73	
Middle Frontal Gyrus	L	−24	41	−6	11		4.3	
Calcarine Fissure	L	−30	−60	10	19	556	4.77	0.014
Calcarine Fissure	L	−21	−55	12	17		3.86	
Calcarine Fissure	L	−15	−60	16	17		3.81	
Superior Temporal Gyrus	L	−44	−13	−3	48	2215	4.71	<0.001
Heschl Gyrus	L	−39	−22	8	48		4.45	
Insula	L	−39	−1	−6	48		4.42	
Rolandic Operculum	R	54	−10	19	48	1621	4.66	<0.001
Rolandic Operculum	R	42	−6	14			4.21	
Insula	R	44	−14	4	48		3.91	
Insula	L	−36	−8	18	48	1019	4.43	<0.001
Rolandic Operculum	L	−38	2	16	48		4.38	
Caudate Nucleus	L	−24	8	25	48		4.36	
Inferior Frontal Gyrus	R	22	35	−6	11	447	4.37	0.036
Insula	R	30	32	10	48		4.17	
Anterior Cingulate Gyrus		14	30	−4	11		3.9	
Caudate Nucleus	R	27	16	20	48	496	4.32	0.024
Anterior Cingulate Gyrus		3	10	22			4.17	
Anterior Cingulate Gyrus		12	16	22			3.45	

Note. Coordinates refer to MNI coordinate system. $p < .05$ cluster-wise FWE corrected (initial cluster-defining threshold $p < .001$). L = left hemisphere, R = right hemisphere.

increased activity during *Look* was found in the left STG, right supramarginal gyrus, and right insula as well as the cingulate gyrus, the precuneus, and the right postcentral gyrus (**Fig. 2E**). Finally, during the third session the control condition resulted in an enhancement of activity in the IFG and insulae, bilaterally, as well as in the occipital cortex (**Fig. 2F**).

3.2.2. Reliability of fMRI results

3.2.2.1. Region-wise reliability. First, we computed ICCs for the mean activation of each ROI within each network to determine region-wise reliability across sessions. In the following, the term "short-term" refers to the comparison between sessions separated by one week (i.e. session 1 | session 2 and session 2 | session 3) while the term "long-term" refers to the longer time interval and the comparison between sessions separated by two weeks (i.e. session 1 | session 3). According to guidelines suggested by Cicchetti and Sparrow (1981) ICCs lower than 0.4 reflect poor reliability (indicated in orange in **Fig. 3**), ICCs between 0.4 and 0.59 fair reliability (indicated in yellow in **Fig. 3**), ICCs between 0.6 and 0.75 good (indicated in light green in **Fig. 3**) and ICCs higher than 0.75 excellent reliability (indicated in green in **Fig. 3**). In fact, according to a recent meta-analysis, most test-retest reliability analyses of task-induced fMRI activity revealed ICC values in the range of poor reliability (56% of the studies), whereas 24% of the studies showed ICC values in the fair and 20% in the good or excellent range of reliability (Elliott et al., 2020). In the following, percentages refer to the number of regions in relation to all regions within one network that lay within a certain range, e.g. fair to good reliability for 40% of the N1 ROIs means that 4 out of 10 regions within N1 showed ICC values within the fair to good range.

N1

Difference contrast. The left posterior MFG, right anterior MFG, right cingulate gyrus (CG), and the precuneus resulted in fair to good ICC values for short- and long-term reliability (40% of the N1 ROIs). Additionally, the left SFG and right insula showed ICC values in the fair to good range, regarding short-term reliability (**Fig. 3A**). Overall, test retest-reliability was higher for the short-term intervals than for the long-term interval, indicated by a higher number of regions exceeding the threshold of fair reliability. The highest ICC values within N1 for short- and long-term reliability were found for prefrontal regions, left posterior and right anterior MFG, respectively.

Main Decrease contrast. ICC values indicated fair to excellent short- and long-term reliability for the majority of ROIs (80% of the N1 ROIs) including prefrontal regions, right CG, the insula, and the precuneus, whereas ICC values of parietal regions of this network (bilateral IPL) were lower and showed only poor to good reliability (Supplementary Material, **Fig. S2A**). The highest ICC values regarding short- and long-term intervals were found for prefrontal regions (left SFG and MFG) and the insula.

Main Look contrast. ICC analyses revealed fair to excellent short-term reliability for the majority of regions (90% of the ROIs within N1; Supplementary Material, **Fig. S3A**). 50% of the regions further exhibited fair to good long-term reliability, including frontal regions (left SFG, right MFG, and CG) and the right insula, indicating a slight decline of reliability with the longer test-retest interval. The right CG and the right insula demonstrated the highest ICC values across each comparison between sessions.

N2

Difference contrast. In general, ICC values were in the fair to excellent range for 88% of the ROIs within N2 with regard to short-term reliability: left posterior SFG, left MFG, left IFG, right IFG, left STG, left MTG, and left caudate (**Fig. 3B**). Similar to the reliability pattern within N1, reliability declined with the longer test-retest interval. In terms of long-term reliability, only 38% of the regions showed ICC values in the fair to excellent range, including left posterior SFG, MFG, and MTG. Again, the highest ICC values within N2 were observed for frontal (left posterior SFG and MFG) and temporal (left MTG) regions.

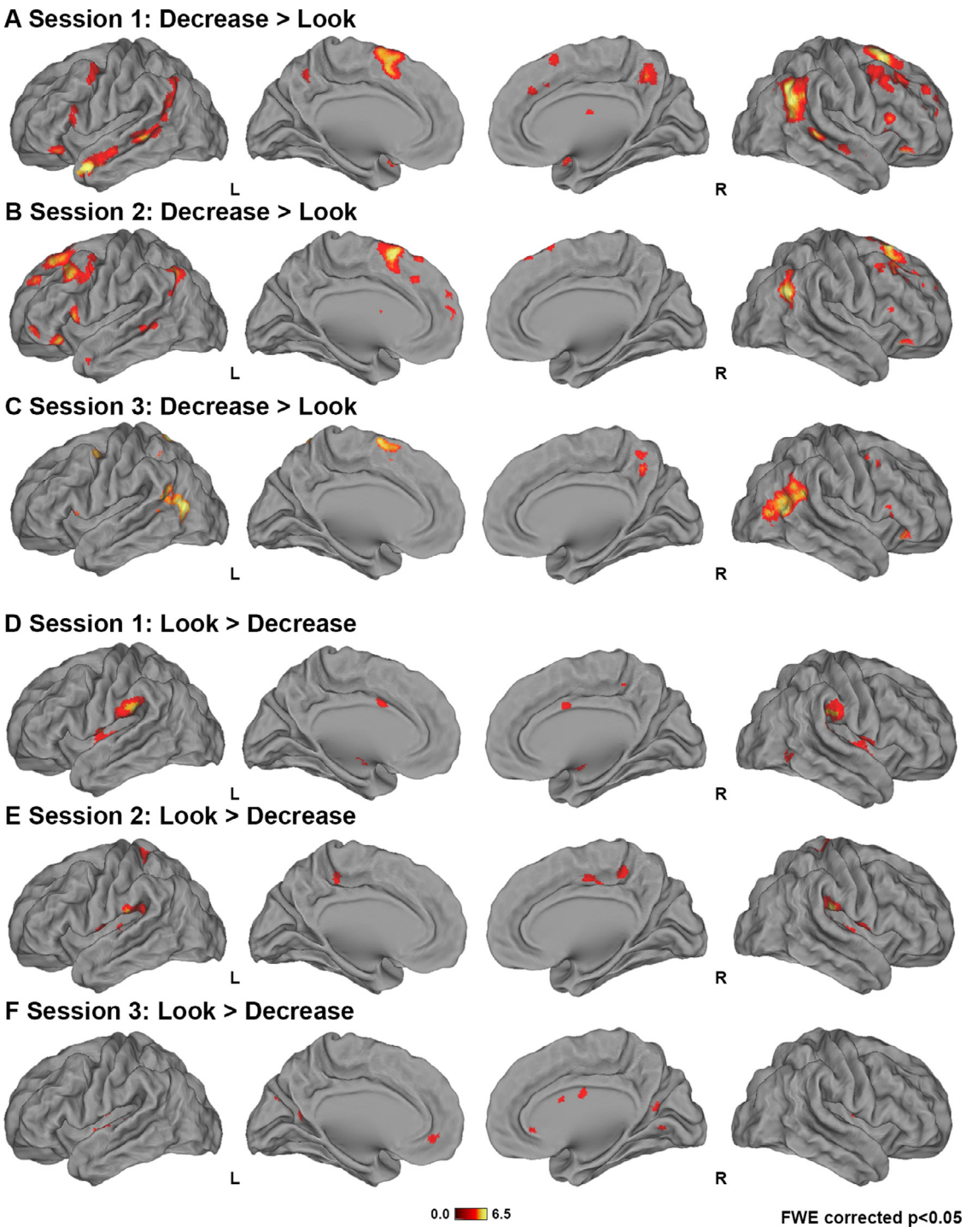

Fig. 2. Brain activation during emotion regulation [*Decrease* > *Look*] for each session separately (**A-C**) and during emotion generation [*Look* > *Decrease*] (**D-F**). *L* = left hemisphere; *R* = right hemisphere.

Main Decrease contrast. ICC values for all regions and time intervals exceeded the threshold of fair short- and long-term reliability (Supplementary Material, **Fig. S2B**). Prefrontal regions, including left posterior SFG, left MFG and bilateral IFG yielded the highest ICC values, revealing good to excellent short- and long-term reliability.

Main Look contrast. Short-term reliability of 38% of the regions within N2 (left posterior SFG, right IFG, and left STG) resulted in fair to good ICC values (Supplementary Material, **Fig. S3B**). Of note, the left posterior SFG also showed good long-term reliability.

N3

Difference contrast. Overall, fair short-term reliability was found for 29% of the regions within N3 including the left parahippocampal gyrus (PHG) as well as the right thalamus across all three sessions (**Fig. 3C**). Only one region, namely the right thalamus, showed ICC values within the fair range for each comparison between sessions.

Main Decrease contrast. Short-term reliability was in the fair to excellent range for 86% of the ROIs (Supplementary Material, **Fig. S2C**). Further, 57% of the regions, including bilateral fusiform gyrus (FG), left

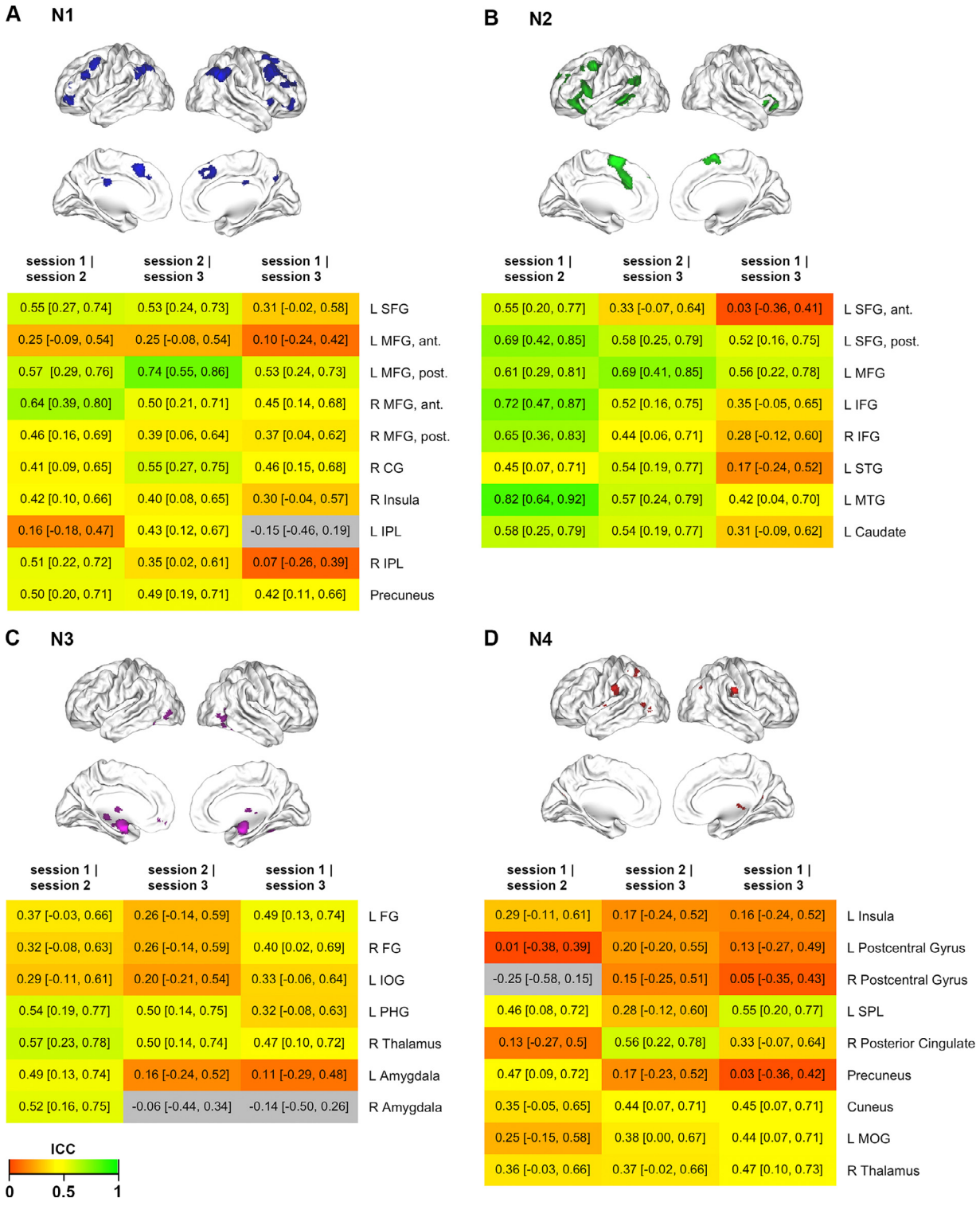

Fig. 3. Heatmaps of region-wise ICCs for ROIs of the four networks (N1 – N4) for the difference contrast. For each network, the ROIs of the respective network (Morawetz et al., 2020) are presented on top. ICC values are shown for each combination of session (short-term intervals: session1 | session2 and session 2 | session 3, long-term interval: session 1 | session 3) with their corresponding confidence interval [CI]. ICC values below 0.40 are considered as low, from 0.4 to 0.59 as fair, from 0.6 to 0.75 as good and above 0.75 as excellent (Cicchetti and Sparrow, 1981). Negative values cannot be interpreted (indicated in gray). Heatmaps of the main contrasts are reported in the Supplementary Material (**Fig. S2** and **S3**).

inferior occipital gyrus (IOG), and left PHG, showed also good long-term reliability. Activity in the FG was related to the highest ICC values in terms of short- and long-term reliability.

Main Look contrast. 43% of the ROIs within N3 showed fair to excellent short-term reliability, namely bilateral FG and left IOG (Supplementary Material, **Fig. S3C**). Further, ICC values of these regions exceeded the excellent range regarding long-term reliability.

N4

Difference contrast. Within N4 not a single region exceeded the threshold of fair reliability for the short-term intervals (**Fig. 3D**).

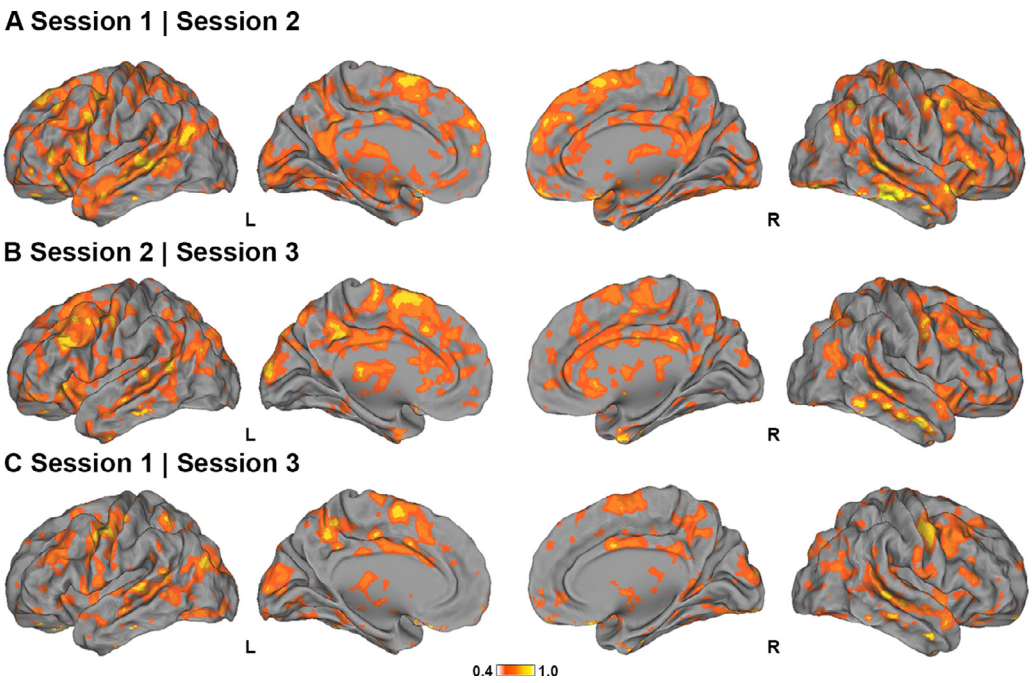

Fig. 4. Whole-brain voxel-wise ICC(3,1) maps for each combination of sessions: session 1 | session 2 (**A**), session 2 | session 3 (**B**), session 1 | session 3 (**C**). ICC values are shown thresholded above 0.40 (fair to excellent reliability).

Main Decrease contrast. Regarding short-term reliability, ICC values were in the fair to excellent range for all ROIs of N4 (Supplementary Material, **Fig. S2D**). Additionally, 67% of the regions were associated with good long-term reliability. The highest ICC values were found for medial parietal regions including the precuneus as well as the posterior cingulate.

Main Look contrast. For the short-term interval, ICC values ranged from fair to excellent for 67% of the ROIs (Supplementary Material, **Fig. S3D**). These regions also showed fair to excellent long-term reliability. The precuneus, left middle occipital gyrus (MOG), and left superior parietal lobe (SPL) tended to have the highest ICC values.

To summarize, the difference contrast yielded fair to excellent reliability for the majority of ROIs in N1 and N2, regarding short-term reliability. Reliability declined within both networks with the longer time interval, indicated by only a few regions that exceeded the threshold of fair long-term reliability. Prefrontal and temporal regions tended to have the highest ICC values within N1 and N2, indicating fair to excellent short- and long-term test-retest reliability. In contrast, overall N3 and N4 showed only poor to fair short- and long-term reliability.

Notably, reliability analyses revealed higher ICC values for the main contrasts (Supplementary Material, **Fig. S2** and **S3**) in comparison to the difference contrast (**Fig. 3**), which is in line with recent findings addressing the issue of reliability of difference contrasts (Infantolino et al., 2018). By using difference score measures, shared reliable variance between the conditions is removed, which results in lower reliability of the difference contrast in comparison to the main contrasts.

3.2.2.2. Voxel-wise reliability. Next, we calculated voxel-wise ICC maps to assess whole-brain voxel-wise reliability (**Fig. 4**). To investigate voxel-wise reliability within each region of each network, we extracted the ICC distributions of each ROI (**Fig. 5**). A reliability measure for each ROI was obtained from the median of the ICC distributions (Caceres et al., 2009) (**Table 3**).

N1

Difference contrast. Within N1, prefrontal regions, including left SFG, bilateral posterior MFG, as well as the precuneus yielded higher ICC values than parietal regions and the insula (**Fig. 5A**). The ICC medians of the former regions showed fair to good short-term reliability (40% of regions within N1) (**Table 3**). Fair long-term reliability was observed for bilateral posterior MFG as well as the precuneus (30% of the N1 ROIs). For the latter regions, voxel-wise ICC values tended to be more dispersed, indicating high variability between ICC values within these regions, and the ICC medians indicated only poor to fair test-retest reliability across the three sessions.

Main Decrease contrast. The ICC distributions for each ROI and each combination of sessions indicated fair to excellent short- and long-term reliability within the whole network (Supplementary Material, **Fig. S4A** and **Table S3**). Prefrontal regions, including left SFG and posterior MFG, yielded the highest ICC values that showed narrow distributions with peaks in the good to excellent range.

Main Look contrast. ICC medians for all regions of N1 were in the fair to good range for the short- and long-term intervals (Supplementary Material, **Table S4**). The highest ICC values were found for right posterior MFG, CG, and insula (Supplementary Material, **Fig. S5A**).

N2

Difference contrast. Regarding short-term reliability, the ICC distributions of the ROIs within N2 showed a high amount of voxel-wise ICC values above the threshold of fair reliability (**Fig. 5B**). This was accompanied by ICC medians in the fair to good range for 88% of the ROIs (**Table 3**). In line with the results of the region-wise ICC analyses, reliability was decreased for the long-term interval, which was reflected in wider ICC distributions that were mainly in the poor range. Only left posterior SFG, MFG, and MTG, resulting in 38% of the regions within the network, yielded ICC medians that revealed fair long-term reliability.

Main Decrease contrast. Medians of the ICC distributions were in the fair to excellent range for all regions within N2, regarding short-term as well as long-term reliability (Supplementary Material, **Table S3**). Note, that the ICC distributions of long-term reliability were wider than the ICC distribution of short-term intervals, indicating more variability of long-term ICC values (Supplementary Material, **Fig. S4B**). Left posterior SFG and MFG as well as bilateral IFG showed the highest ICC values across each combination of sessions followed by the temporal regions (left STG and MTG) of the network.

Main Look contrast. 75% of regions (bilateral SFG, left MFG, right IFG, left STG, and MTG) showed fair to good short- and long-term reliability

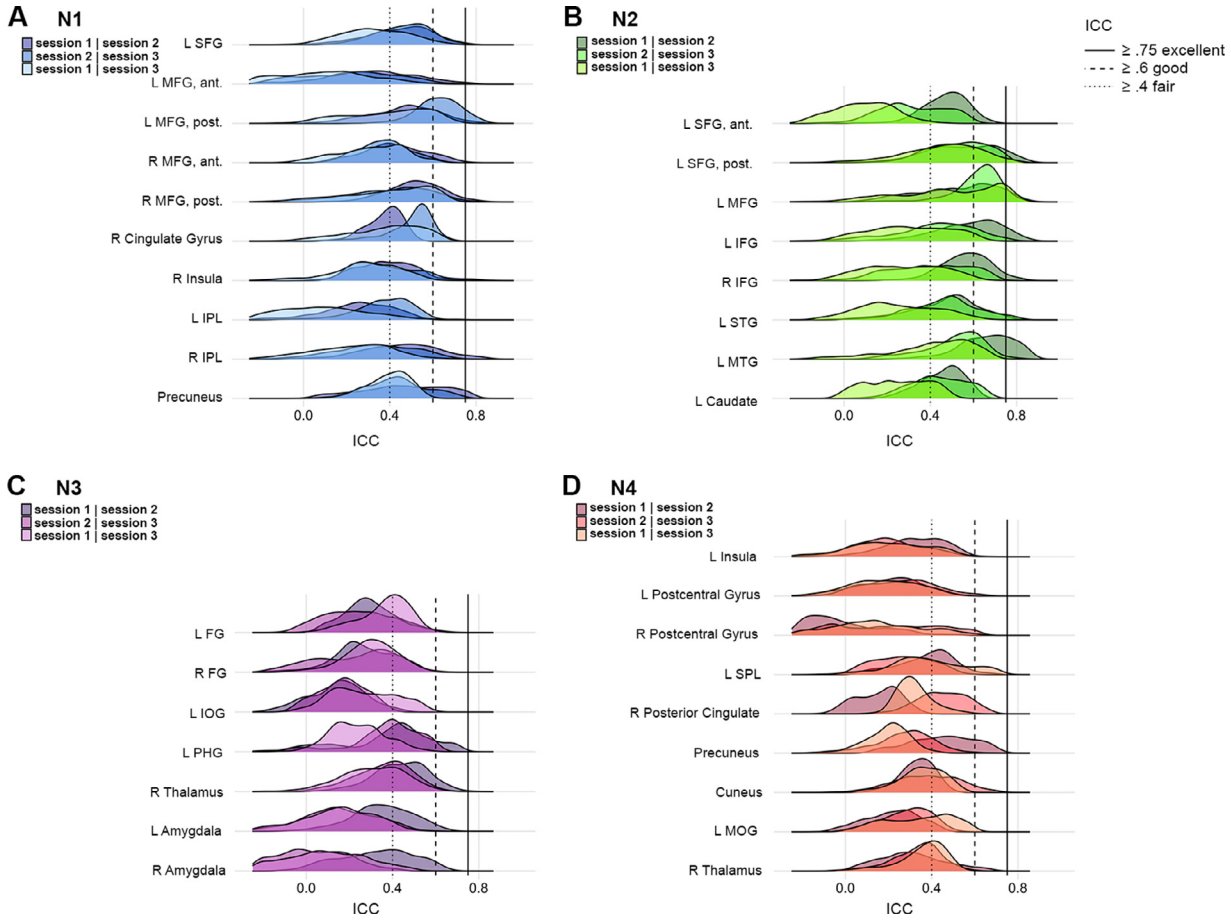

Fig. 5. Density plots of voxel-wise ICCs for ROIs of the four networks (N1- N4) for the difference contrast. For each network and each respective ROI, the distribution of voxel-wise ICC values is presented for each combination of session (one-week intervals: session1 | session2 and session 2 | session 3, two-week interval: session 1 | session 3). Horizontal lines indicate the categorization of ICC values as introduced by Cicchetti and Sparrow (1981): ICC values below 0.40 (dotted line) are considered as low, from 0.4 to 0.59 as fair, from 0.6 (dashed line) to 0.75 (continuous line) as good and above 0.75 as excellent. Density plots of the main contrasts are reported in the Supplementary Material (**Fig. S4** and **S5**).

(Supplementary Material, **Fig. S5B** and **Table S4**). Reliability declined with the longer time interval as indicated by a shift of the long-term ICC distribution to lower values. The highest ICC values were observed for prefrontal regions, including left SFG and MFG.

N3

Difference contrast. The ICC distributions of the ROIs within N3 showed a high amount of voxel-wise ICC values below the threshold of fair reliability (**Fig. 5C**) as indicated by the medians of the ICC distributions, fair short-term reliability was only found for one region within N3, namely the left PHG (**Table 3**) No region exceeded the threshold of fair long-term reliability.

Main Decrease contrast. 57% of the regions (bilateral FG, left IOG, and left PHG) showed fair to excellent ICC medians for each session interval (Supplementary Material, **Table S3**). The highest ICC values were found for bilateral FG (Supplementary Material, **Fig. S4C**), which was further supported by ICC medians in the excellent short- and long-term range. Note, that the medians of the ICC distributions for the bilateral amygdala (and left PHG) were above the fair range in terms of short-term reliability. However, the broad ICC distributions values indicated high variability of voxel-wise ICC values within these regions and are therefore difficult to interpret regarding the reliability of the ROIs.

Main Look contrast. Only 43% of the regions exceeded the fair reliability threshold for short- and long-term reliability (Supplementary Material, **Table S4**). The ICC distributions and the corresponding ICC medians revealed good to excellent reliability for bilateral FG and left IOG (Supplementary Material, **Fig. S5C** and **Table S4**). The subcortical regions of the network showed only poor reliability and high variability of the ICC values.

N4

Difference contrast. Within N4, a high amount of the voxel-wise ICC values fell in the poor range of reliability for each ROI (**Fig. 5D**). This was reflected by the medians of the ICC distributions (**Table 3**): No region survived the fair short- or long-term reliability threshold, indicating poor test-retest reliability within the whole network.

Main Decrease contrast. 56% of the regions (left SPL, right posterior cingulate, the precuneus, the cuneus, and left MOG) showed good to excellent short- and long-term reliability as indicated by the ICC medians (Supplementary Material, **Table S3**) as well as their ICC distributions that lied above the threshold of good reliability (Supplementary Material, **Fig. S4D**). The remaining regions yielded also ICC medians in the fair to good range for the short-term (left insula, bilateral postcentral gyrus, and right thalamus) and long-term (left postcentral gyrus, right thalamus) interval, but their distributions showed high variability of voxel-wise ICC values.

Main Look contrast. ICC medians revealed fair to excellent short- and long-term reliability for 78% of the regions within the network (left postcentral gyrus, left SPL, right posterior cingulate, the precuneus, the cuneus, left MOG, and right thalamus) (Supplementary Material, **Table S4**). However, voxel-wise ICC values of long-term reliability for the right thalamus vary substantially (Supplementary Material, **Fig. S5D**).

To summarize, we found a similar pattern of *voxel-wise* reliability as in the *region-wise* reliability approach, which is in line with previous

Table 3
Medians of voxel-wise ICC distributions for each ROI.

Region	ICC Median		
	Session 1 | Session 2	Session 2 | Session 3	Session 1 | Session 3
N1			
L SFG	0.46	0.47	0.33
L MFG, ant.	0.31	0.24	0.09
L MFG, post.	0.48	0.63	0.43
R MFG, ant.	0.41	0.38	0.33
R MFG, post.	0.52	0.42	0.43
R CG	0.39	0.53	0.41
R Insula	0.38	0.36	0.30
L IPL	0.26	0.39	0.07
R IPL	0.46	0.33	0.24
Precuneus	0.47	0.43	0.41
N2			
L SFG, ant.	0.48	0.31	0.10
L SFG, post.	0.55	0.54	0.47
L MFG	0.55	0.64	0.51
L IFG	0.60	0.44	0.30
R IFG	0.57	0.40	0.27
L STG	0.49	0.48	0.24
L MTG	0.67	0.54	0.44
L Caudate	0.47	0.44	0.24
N3			
L FG	0.29	0.23	0.39
R FG	0.26	0.26	0.30
L IOG	0.16	0.17	0.25
L PHG	0.42	0.40	0.23
R Thalamus	0.45	0.37	0.32
L Amygdala	0.35	0.16	0.15
R Amygdala	0.34	–0.03	0.03
N4			
L Insula	0.33	0.19	0.20
L Postcentral Gyrus	0.22	0.26	0.18
R Postcentral Gyrus	–0.07	0.17	0.11
L SPL	0.40	0.27	0.37
R Posterior Cingulate	0.16	0.45	0.31
Precuneus	0.46	0.29	0.21
Cuneus	0.34	0.38	0.36
L MOG	0.22	0.30	0.36
R Thalamus	0.30	0.35	0.38

Note. ROIs within the same brain regions are further labeled as the anterior (ant.) or posterior (post.) ROI, respectively. L = left hemisphere, R = right hemisphere.

findings (Fröhner et al., 2019). For the difference contrast, the ICC distributions and their corresponding ICC medians indicated the highest test-retest reliability for prefrontal (MFG and SFG) as well as temporal regions (left MTG) that are part of the emotion regulation networks (N1 and N2). Corresponding to the *region-wise* ICC analyses, reliability declined within both networks with the longer time interval (**Fig. 4**), which was reflected by lower ICC medians and high variability of ICC values within the respective ROIs. Regions within N3 and N4 showed only poor short- and long-term reliability on a voxel level.

Again, the reliability analyses using the main contrasts revealed higher *voxel-wise* ICC values for the majority of ROIs (Supplementary Material, **Fig. S4** and **S5**) in comparison to the difference contrast (**Fig. 5**).

4. Discussion

Here we addressed for the first time the issue of the test-retest reliability of brain activation during emotion regulation by acquiring 7T fMRI data at three sessions separated by one week. Specifically, we focused on the reliability of regions within four different networks shown to be involved in emotion generation and regulation that were derived from a recent meta-analysis (Morawetz et al., 2020). The effect of emotion regulation was demonstrated on the behavioral and neural level. First, engaging in emotion regulation resulted in decreased negative emotional state ratings compared to the control condition within each session. Second, in accord with previous findings, emotion regulation compared to the control condition across all sessions was associated with increased activation in frontal (IFG, SFG, SMA), temporal (MTG), and parietal (precuneus) regions (e.g., Kohn et al., 2014; Morawetz et al., 2017). During emotion generation, i.e. control condition vs. emotion regulation condition, activity in the frontal (IFG, SMA), the temporal (STG) and the cingulate cortex as well as the insula was enhanced, in line with previous research (e.g., Morawetz et al., 2017; Ochsner et al., 2004). With regard to test-retest reliability, our results yielded good reliability of a well-established emotion regulation task on a behavioral level. Test-retest reliability of underlying neural networks varied considerably across the networks and respective ROIs. Importantly, prefrontal and temporal regions demonstrated good to excellent test-retest reliability during the down-regulation of emotions, which implies that these regions might represent stable core regions supporting cognitive emotion regulation that can be studied on an individual subject level.

The whole-brain analyses on a group level revealed that brain activity was relatively stable across the three sessions in regions involved in emotion generation and regulation. However, when comparing the group level responses between sessions, we observed a decrease in activation for the emotion generation and regulation contrasts over time. A number of studies have reported a reduction of brain activation for subcortical and cortical regions during emotion processing

(Geissberger et al., 2020; Lipp et al., 2014; Sauder et al., 2014) and cognitive tasks (Plichta et al., 2012) over multiple time points. Notably, to avoid habituation effects that have been reported for the repeated presentation of negative stimuli in prefrontal cortex regions, the anterior cingulate gyrus and the amygdala (Denny et al., 2014; Denny and Ochsner, 2015; Phan et al., 2003; Wright et al., 2001), we optimized our study design by using different images for each session and condition, that were matched for content and stimulus features. Habituation effects to the stimuli can therefore be ruled out. One alternative explanation might be a decline in task-related motivation and engagement over time resulting in decreased brain responses. However, our behavioral findings only revealed an effect of session within the *Look* condition, i.e. less negative ratings within the *Look* condition over time, while emotional state ratings within the *Decrease* condition remained relatively stable across sessions. Thus, this alternative explanation of a decline in task-related motivation and engagement becomes unlikely.

This directly leads to a third explanation. Given the stability of the emotional state ratings, the decrease in activation on a whole-brain level might correspond to enhanced neural efficiency in emotion regulatory processes over time (e.g., Ramsey et al., 2004; Schweizer et al., 2013). Neural efficiency is characterized by decreased task-related neural activation accompanied by greater or stable task performance (Gray et al., 2005; Kelly and Garavan, 2005). This decrease in neural activation reflects increased efficiency of the underlying neural circuits and might be related to lower cognitive effort (Neubauer and Fink, 2009). The idea of neural efficiency fits our data, as neural activation during emotion regulation decreased across sessions. However, this interpretation of our results has to be treated with caution, as we did not implement any explicit emotion regulation training, that could explain a training effect resulting in higher neural efficiency and less cognitive effort. Further, in our task design, we did not acquire an objective measure of cognitive effort, e.g. pupil dilatation (Maier and Grueschow, 2020) and therefore cannot provide additional evidence in support of the neural efficiency model by demonstrating less effortful cognitive processing over sessions.

With regard to test-retest reliability, we demonstrate region-wise and voxel-wise reliability of brain activation during emotion regulation that ranged from poor to excellent and was further corroborated by good to excellent reliability of emotional state ratings. As the consistency of behavioral performance is associated with test-retest reliability of brain responses, the good test-retest reliability of our behavioral measures has to be emphasized (Fröhner et al., 2019). Our results suggest, that emotional state ratings as well as regulation success provide a reliable measure for the assessment of differences between individuals on a behavioral level. On a neural level, the comparison of emotion generation and regulation networks revealed opposed reliability patterns. While ICC analyses of emotion regulatory networks (N1 and N2) mainly based on prefrontal, parietal and temporal regions yielded fair to excellent reliability, emotion generative networks (N3 and N4) remained in the poor to fair range of reliability estimates. Several factors might contribute to this finding. First, the location of the regions might play an important role. Whereas N1 and N2 include mainly cortical regions, i.e. frontal, temporal and parietal regions, N3 and N4 also include subcortical regions, such as the amygdala, caudate and thalamus. This is in line with a recent meta-analysis reporting that cortical ROIs show significantly higher reliability estimates than subcortical regions (Elliott et al., 2020). One explanation for lower ICCs in subcortical regions might be the proneness of ventral brain regions to physiological noise and susceptibility artefacts due to their proximity to bone and sinuses (Merboldt et al., 2001; Robinson et al., 2004). At ultra-high field strength, susceptibility effects as well as physiological artefacts become more severe (Moser et al., 2012; Triantafyllou et al., 2005) and thus might have caused the observed regional differences in test-retest reliability. However, optimized scanning protocols including high spatial resolution lead to a reduction of the relative field dispersions within a voxel and thus reduces dephasing and signal dropout (Balchandani and Naidich, 2015; Sladky et al., 2013; Triantafyllou et al., 2005). As demonstrated in several previous studies, high spatial resolution approaches as used in our study effectively compensate for increased signal losses from susceptibility artefacts (e.g., Geissberger et al., 2020; Hahn et al., 2013; Robinson et al., 2004; Sladky et al., 2013). Another possible explanation for differences in test-retest-reliability measures between the networks might be the magnitude of activation within the ROIs. It has been demonstrated that higher levels of activation are related to higher ICC values (Caceres et al., 2009; Fliessbach et al., 2010; Korucuoglu et al., 2020). Thus, the fact that brain regions within N1 and N2, in particular prefrontal and temporal regions, were associated with higher activity than regions within N3 might explain the overall higher ICC values. This is due to the task design and resulting contrasts as the engagement in the active task of regulation clearly induced the heightened activity in N1 and N2 regions, while the passive viewing condition naturally resulted in less recruitment of N3 regions and poor reliability. Overall, our findings suggest, that regions within N1 and N2 can be reliably measured by conducting an emotion regulation task whereas the same task might be less applicable to measure emotion generative processes on an individual subject level.

As most of the regions within N1 and N2 turned out to be at least fairly reliable in the region-wise as well as the voxel-wise analyses, we further focus our discussion on these networks to identify core regions implicated in emotion regulation. The highest ICC values were found for prefrontal and temporal regions, including left posterior MFG/dlPFC, right anterior MFG/vlPFC, left posterior SFG/SMA, and left MTG. These regions seem to play a key role within the emotion regulatory process being involved in attention and working memory (dlPFC), in cognitive control and the selection of goal-appropriate responses (vlPFC), in the formation of mental representations (SMA), and in perceptional and semantic emotional representations (MTG) (Kohn et al., 2014; Ochsner et al., 2012; Silvers and Guassi Moreira, 2019). Our findings suggest that these prefrontal and temporal regions might serve as promising candidates for the study of individual differences in emotion regulation, e.g. the interaction of emotion regulatory processes with age (e.g., Silvers et al., 2017), personality (e.g., S. Chen et al., 2017) or habitual use of emotion regulation strategies (e.g., Abler et al., 2010; Vanderhasselt et al., 2013) and for neurobiological biomarkers in a clinical context. A large body of research has shown that alterations in emotion regulation are associated with various psychopathologies thereby describing emotion regulation as a transdiagnostic phenomenon (Cludius et al., 2020). In line with this view, reduced recruitment of regions within the emotion regulatory network, including the vlPFC and dlPFC (which demonstrate the highest reliability in our study), has been associated with emotional dysregulation in a range of psychological disorders (i.e., mood disorders, anxiety disorders, addiction, schizophrenia and personality disorders) (Picó-Pérez et al., 2017; Zilverstand et al., 2017). As a result, emotion regulation and cognitive control serve as targets for behavioral, pharmacological and neuromodulatory interventions (Cludius et al., 2020; Cohen and Ochsner, 2018; Paret and Hendler, 2020; Roiser et al., 2012). Investigating the neural correlates of emotion regulation interventions by using fMRI provides a useful tool to study the interaction of person factors (e.g., age, personality traits or severity of psychopathological symptoms) and treatment mechanisms (Denny, 2020), and to provide predictive biomarkers of treatment response (Roiser et al., 2012). Furthermore, several studies suggest, that abnormal activation patterns in the regulatory network might serve as prognostic biomarkers for future psychological diseases (e.g., Heissler et al., 2014; Kanske et al., 2012; Van Der Velde et al., 2015). To build such predictive models that focus on clinical outcomes on the one hand and on basic neural processes that might be dysregulated across multiple disorders on the other hand, the diagnostic value of biomarkers needs to be assessed through different methods (Woo et al., 2017). Test-retest reliability represents one of these methods and thus, must be at the forefront of study design in order to develop predictive models in the clinical context that are

applicable to individual persons and neuroscientifically plausible and interpretable.

Notably, to the best of our knowledge, this is the first study on emotion regulation conducted at ultra-high magnetic field (7T). Due to the promising benefits of ultra-high field strength, ultra-high field MR constitutes a powerful tool within the field of clinical and research-related neuroimaging (Balchandani and Naidich, 2015; Morris et al., 2019; Moser et al., 2012). Ultra-high field MR of human brain structures and function has become increasingly available. Higher magnetic fields offer enhanced sensitivity by effectively increasing image and time course signal-to-noise ratio, spatial resolution, and sensitivity to susceptibility effects resulting in higher BOLD-related signal changes (Balchandani and Naidich, 2015; Moser et al., 2012; Uğurbil et al., 2003). By increasing signal-to-noise and contrast-to-noise ratios, ultra-high field MRI provides a way for improving test-retest reliability for fMRI (Bennett and Miller, 2010), and constitutes a promising tool for the investigation of individual difference. A recent study (Geissberger et al., 2020) suggested, that the test-retest reliability of task fMRI can benefit from ultra-high field strength. Using high spatial resolution acquisition at ultra-high field they effectively compensated for increased signal losses from susceptibility artefacts and demonstrated generally higher ICC values (within the range of excellent test-retest reliability) of neural activation than previous studies using lower field strength for a well-studied emotion processing paradigm. Our findings suggest that fMRI at ultra-high field provides a promising method for investigating emotion regulation on an individual level by studying brain activation in core regions for emotion regulatory processes.

This study has several limitations. It has to be noted that the data acquisition at ultra-high magnetic field (7T) limits the generalizability of our results. As fMRI studies exploring emotion regulation typically use fMRI at high magnetic field (3T), test-retest reliability of brain activity during emotion regulation using 3T fMRI needs to be addressed in future studies. However, we believe that our study can inform future fMRI research on emotion regulation as the number of fMRI studies at ultra-high field strength is increasing and, more importantly, we investigated test-retest reliability of brain activity underlying emotion regulation for the first time. Further, our results are dependent on our study sample primarily consisting of young females, which might limit the generalizability to other populations. Therefore, future studies should include larger and more diverse samples.

In addition, methodological issues regarding our study design might be considered in future studies. Firstly, in our study stimuli were not counterbalanced across sessions and participants. Secondly, we further acknowledge that our test-retest reliability results are dependent on our study design and protocol, including the one-week time interval between sessions. Thirdly, our study only implemented reappraisal as emotion regulation strategy. Finally, we did not assess current mood in each session and thus cannot completely rule out that changes in mood might have affected emotional reactivity and regulation of the participants. This might have led to increased within-subject variability across sessions resulting in decreased test-retest reliability. However, the high test-retest reliability of the emotional state ratings does not support the idea of mood-related effects. Taken together, future studies could extend our findings by (a) incorporating a counterbalanced design, (b) investigating longer test-retest intervals in longitudinal studies (e.g. several weeks to months), (c) implementing variations of the task (e.g. different emotion regulation strategies), and (d) considering state-dependent psychological processes such as mood.

It is worth mentioning, that we found higher reliability estimates within each network for the main contrasts of the conditions, i.e. the *Decrease* and the *Look* condition, compared to the difference contrast. This finding is not surprising, as change scores, i.e. the difference of two measures, are less reliable than the measures themselves when these measures are correlated (Hedge et al., 2018) resulting in lower reliability of difference contrasts compared to the constituent main contrasts (Infantolino et al., 2018). Thus, when examining differences between individuals in emotion regulation, one possibility to enhance the reliability of the assessment of neural correlates might be the extraction of the beta values based on the main contrasts for individual difference analyses. Another possibility to ensure good reliability could be the inclusion of an explicit baseline condition (neutral stimuli) in the task that might offer more reliable change scores, as a baseline condition and the regulation condition might share less between-subject variance that is canceled out by the subtraction of the scores. Furthermore, recent reviews on fMRI reliability emphasize the advantage of multivariate fMRI analyses in terms of higher test-retest reliability compared to univariate measures (Kragel et al., 2020; Noble et al., 2020). Thus, using a multivariate approach (e.g., (Morawetz et al., 2021; Morawetz et al., 2016; Powers et al., 2020) might further enhance test-retest reliability and therefore suitability for studying individual differences in emotion regulation.

Of note, we found no significant activation in the amygdala for the emotion generation contrast on the whole-brain level, despite it being commonly reported to be modulated by emotion regulation in previous studies (e.g., Dörfel et al., 2014; Kanske et al., 2011; Ochsner et al., 2004). However, there are also numerous studies that failed to find a modulating effect of emotion regulation on amygdala activity on a whole-brain level (e.g., McRae et al., 2012; Morawetz et al., 2016; Otto et al., 2014; Silvers et al., 2015). These controversial findings could be explained by methodological issues such as stimulus features (e.g. arousal) (Dolcos et al., 2014), habituation effects (Geissberger et al., 2020), or proneness to susceptibility artifacts (Merboldt et al., 2001; Morawetz et al., 2008; Robinson et al., 2004). The latter reason could be ruled out in our study, given the high resolution at ultra-high magnetic fields (Geissberger et al., 2020; Sladky et al., 2013). Thus, the lack of amygdala activity during the control condition might be explained by the fact that we used a wide range of low and high arousing stimuli, canceling any effects due to different levels of arousal.

5. Conclusions

In this study, we examined for the first time the test-retest reliability of brain activity during the cognitive control of emotions using reappraisal and fMRI. Our findings suggest, that the reliability of brain activation is good to excellent in core emotion regulatory regions, which has several implications: First, the task investigated in our study is highly applicable to identify individual differences in emotion regulation performance. Second, while the measurement of neural activation underlying emotion generative processes on an individual level has to be treated with caution, we identified several regions (vlPFC, dlPFC, MTG) that showed good test-retest reliability for the assessment of emotion regulative processes. These regions therefore might qualify as good candidates for the study of individual differences in emotion regulation as well as for biomarkers in clinical neuroscience research. In light of the ongoing reliability discussion, our results are of high relevance and interest for fMRI research in general and emotion regulation research in particular, as we provide the first step toward the investigation of reliability of fMRI measures that are widely applied within the affective neuroscience.

Data statement

The data that support the findings of this study are available from the corresponding author, C.M., upon reasonable request.

Author statement

Stella Berboth: Methodology, Software, Validation, Formal Analysis, Investigation, Writing – Original Draft, Writing – Review & Editing, Visualization. **Christian Windischberger**: Conceptualization, Methodology, Resources, Writing – Original Draft, Writing – Review & Editing,

Project administration, Funding acquisition. **Nils Kohn**: Conzeptualization, Methodology, Formal Analysis, Writing – Original Draft, Writing – Review & Editing. **Carmen Morawetz:** Conzeptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Resources, Writing – Original Draft, Writing – Review & Editing, Visualization, Supervision, Project administration, Funding acquisition.

Funding

This work was supported by Marie Skłodowska-Curie Action 795994 to C.M.

Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

We would like to thank Alexandra Mindu and Lucas Jeay-Bizot for help with the data collection.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2021.117917.

References

Abler, B., Hofer, C., Walter, H., Erk, S., Hoffmann, H., Traue, H.C., Kessler, H., 2010. Habitual emotion regulation strategies and depressive symptoms in healthy subjects predict fMRI brain activation patterns related to major depression. Psych. Res. - Neuroimag. 183 (2), 105–113. doi:10.1016/j.pscychresns.2010.05.010.

Balchandani, P., Naidich, T.P., 2015. Ultra-high-field MR neuroimaging. Am. J. Neuroradiol. 36 (7), 1204–1215. doi:10.3174/ajnr.A4180.

Bartko, J.J., 1976. On various intraclass correlation reliability coefficients. Psychol. Bull. 83 (5), 762–765.

Bennett, C.M., Miller, M.B., 2010. How reliable are the results from functional magnetic resonance imaging ? Ann. N. Y. Acad. Sci. 1191, 133–155. doi:10.1111/j.1749-6632.2010.05446.x.

Berking, M., Wupperman, P., 2012. Emotion regulation and mental health: recent findings, current challenges, and future directions. Curr. Opin. Psychiatry 25 (2), 128–134. doi:10.1097/YCO.0b013e3283503669.

Bradley, M.M., Lang, P.J., 2007. The International affective picture system (IAPS) in the study of emotion and attention. In: Coan, J.A., Allen, J.J. (Eds.), Handbook of Emotion Elicitation and Assessment. Oxford University Press, New York, pp. 29–46.

Brett, M., Anton, J.-.L., Valabregue, R., Poline, J.-.B., 2002. Region of interest analysis using an SPM toolbox [abstract]. Neuroimage 16 (2), 497 Abstract.

Buhle, J.T., Silvers, J.A., Wager, T.D., Lopez, R., Onyemekwu, C., Kober, H., … Ochsner, K.N., 2014. Cognitive reappraisal of emotion: a meta-analysis of human neuroimaging studies. Cereb. Cortex 24 (11), 2981–2990. doi:10.1093/cercor/bht154.

Caceres, A., Hall, D.L., Zelaya, F.O., Williams, S.C.R., Mehta, M.A, 2009. Measuring fMRI reliability with the intra-class correlation coefficient. Neuroimage 45, 758–768 https://doi.org/10.1016/j.neuroimage.2008.12.035.

Chen, G., Taylor, P.A., Haller, S.P., Kircanski, K., Stoddard, J., Pine, D.S., … Cox, R.W., 2018. Intraclass correlation : improved modeling approaches and applications for neuroimaging. Hum. Brain Mapp. 39, 1187–1206. doi:10.1002/hbm.23909.

Chen, S., Chen, C., Yang, J., Yuan, J., 2017. Trait self-consciousness predicts amygdala activation and its functional brain connectivity during emotional suppression: an fMRI analysis. Sci. Rep. 7 (1), 1–11. doi:10.1038/s41598-017-00073-3.

Cicchetti, D.V., Sparrow, S.A., 1981. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. Am. J. Ment. Defic. 86 (2).

Cludius, B., Mennin, D., Ehring, T., 2020. Emotion regulation as a transdiagnostic process. Emotion 20 (1), 37–42. doi:10.1037/emo0000646.

Cohen, N., Ochsner, K.N., 2018. From surviving to thriving in the face of threats: the emerging science of emotion regulation training. Curr. Opin. Behav. Sci. 24, 143–155. doi:10.1016/j.cobeha.2018.08.007.

Denny, B.T., 2020. Getting better over time: a framework for examining the impact of emotion regulation training. Emotion 20 (1), 110–114. doi:10.1037/emo0000641.

Denny, B.T., Fan, J., Liu, X., Guerreri, S., Mayson, S.J., Rimsky, L., … Koenigsberg, H.W., 2014. Insula-amygdala functional connectivity is correlated with habituation to repeated negative images. Soc. Cogn. Affect. Neurosci. 9 (11), 1660–1667. doi:10.1093/scan/nst160.

Denny, B.T., Inhoff, M.C., Zerubavel, N., Davachi, L., Ochsner, K.N., 2015. Getting over it: long-lasting effects of emotion regulation on amygdala response. Psychol. Sci. 26 (9), 1377–1388. doi:10.1177/0956797615578863.

Denny, B.T., Ochsner, K.N., 2015. Behavioral effects of longitudinal training in cognitive reappraisal. Emotion 14 (2), 425–433. doi:10.1037/a0035276.Behavioral.

Dolcos, S., Katsumi, Y., Dixon, R.A., 2014. The role of arousal in the spontaneous regulation of emotions in healthy aging: a fMRI investigation. Front. Psychol. 5, 1–12. doi:10.3389/fpsyg.2014.00681.

Dörfel, D., Lamke, J.P., Hummel, F., Wagner, U., Erk, S., Walter, H., 2014. Common and differential neural networks of emotion regulation by detachment, reinterpretation, distraction, and expressive suppression: a comparative fMRI investigation. Neuroimage 101, 298–309. doi:10.1016/j.neuroimage.2014.06.051.

Dubois, J., Adolphs, R., 2016. Building a science of individual differences from fMRI. Trends Cogn. Sci. 20 (6), 425–443. doi:10.1016/j.tics.2016.03.014.

Eftekhari, A., Zoellner, L.A., Vigil, S.A., 2009. Patterns of emotion regulation and psychopathology. Anxiety Stress Coping 22 (5), 571–586. doi:10.1080/10615800802179860.

Elliott, M., Knodt, A., Ireland, D., Morris, M., Poulton, R., Ramrakha, S., … Hariri, A., 2020. What is the test-retest reliability of common task-fMRI Measures? New empirical evidence and a meta-analysis. Psychol. Sci. 87 (9), 792–806. doi:10.1016/j.biopsych.2020.02.356.

Fliessbach, K., Rohe, T., Linder, N.S., Trautner, P., Elger, C.E., Weber, B., 2010. Retest reliability of reward-related BOLD signals. Neuroimage 50 (3), 1168–1176. doi:10.1016/j.neuroimage.2010.01.036.

Fostering reproducible fMRI research, 2017. Nat. Neurosci. 20 (3), 298. doi:10.1038/nn.4521.

Fröhner, J.H., Teckentrup, V., Smolka, M.N., Kroemer, N.B., 2019. Addressing the reliability fallacy in fMRI: similar group effects may arise from unreliable individual effects. Neuroimage 195, 174–189. doi:10.1016/j.neuroimage.2019.03.053.

Geissberger, N., Tik, M., Sladky, R., Woletz, M., Schuler, A.L., Willinger, D., Windischberger, C., 2020. Reproducibility of amygdala activation in facial emotion processing at 7T. Neuroimage 211, 116585. doi:10.1016/j.neuroimage.2020.116585.

Giradeau, B., 1996. Negative values of the intraclass correlation coefficient are not theoretically possible. J. Clin. Epidemiol. 49 (10), 1205–1206.

Gray, J.R., Burgess, G.C., Schaefer, A., Yarkoni, T., Larsen, R.J., Braver, T.S., 2005. Affective personality differences in neural processing efficiency confirmed using fMRI. Cognit., Affect. Behav. Neurosci. 5 (2), 182–190. doi:10.3758/CABN.5.2.182.

Gross, J.J., John, O.P., 2003. Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. J. Pers. Soc. Psychol. 85 (2), 348–362. doi:10.1037/0022-3514.85.2.348.

Gross, J.J., Sheppes, G., Urry, H.L., 2011. Cognition and emotion lecture at the 2010 SPSP emotion preconference: emotion generation and emotion regulation: a distinction we should make (Carefully). Cognit. Emotion 25 (5), 765–781. doi:10.1080/02699931.2011.555753.

Hahn, A., Kranz, G.S., Seidel, E.M., Sladky, R., Kraus, C., Küblböck, M., … Lanzenberger, R., 2013. Comparing neural response to painful electrical stimulation with functional MRI at 3 and 7T. Neuroimage 82, 336–343. doi:10.1016/j.neuroimage.2013.06.010.

Hedge, C., Powell, G., Sumner, P., 2018. The reliability paradox: why robust cognitive tasks do not produce reliable individual differences. Behav. Res. Methods 50 (3), 1166–1186 https://doi.org/10.3758/s13428-017-0935-1.

Heissler, J., Kanske, P., Schönfelder, S., Wessa, M., 2014. Inefficiency of emotion regulation as vulnerability marker for bipolar disorder: evidence from healthy individuals with hypomanic personality. J. Affect. Disord. 152–154 (1), 83–90. doi:10.1016/j.jad.2013.05.001.

Infantolino, Z.P., Luking, K.R., Sauder, C.L., Curtin, J.J., Hajcak, G., 2018. Robust is not necessarily reliable: from within-subjects fMRI contrasts to between-subjects comparisons. Neuroimage 173, 146–152. doi:10.1016/j.neuroimage.2018.02.024.Robust.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D., Quinn, K., … Wang, P., 2010. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Am. J. Psychiatry 167 (7), 748–751. doi:10.1176/appi.ajp.2010.09091379.

Kanske, P., Heissler, J., Schönfelder, S., Bongers, A., Wessa, M., 2011. How to regulate emotion? Neural networks for reappraisal and distraction. Cereb. Cortex 21 (6), 1379–1388. doi:10.1093/cercor/bhq216.

Kanske, P., Heissler, J., Schönfelder, S., Wessa, M., 2012. Neural correlates of emotion regulation deficits in remitted depression: the influence of regulation strategy, habitual regulation use, and emotional valence. Neuroimage 61 (3), 686–693. doi:10.1016/j.neuroimage.2012.03.089.

Kelly, A.M.C., Garavan, H, 2005. Human functional neuroimaging of brain changes associated with practice. Cereb. Cortex 15 (8), 1089–1102. doi:10.1093/cercor/bhi005.

Kohn, N., Eickhoff, S.B., Scheller, M., Laird, A.R., Fox, P.T., Habel, U., 2014. Neural network of cognitive emotion regulation–an ALE meta-analysis and MACM analysis. Neuroimage 87, 345–355. doi:10.1016/j.neuroimage.2013.11.001.

Koo, T.K., Li, M.Y., 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J. Chiropr. Med. 15, 155–163. doi:10.1016/j.jcm.2016.02.012.

Korucuoglu, O., Harms, M.P., Astafiev, S.V., Kennedy, J.T., Golosheykin, S., Barch, D.M., Anokhin, A.P., 2020. Test-retest reliability of fMRI-measured brain activity during decision making under risk. Neuroimage 214, 116759. doi:10.1016/j.neuroimage.2020.116759.

Kragel, P.A., Han, X., Kraynak, T.E., Gianaros, P.J., & Wager, T.D. (2020). fMRI can be highly reliable, but it depends on what you measure, 1–11.

Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S.F., Baker, C.I, 2009. Circular analysis in systems neuroscience: the dangers of double dipping. Nat. Neurosci. 12 (5), 535–540. doi:10.1038/nn.2303.Circular.

Kring, A.M., Sloan, D.M.X, 2010. Emotion Regulation and psychopathology: A transdiagnostic Approach to Etiology and Treatment. Guilford Press, New York.

Lahey, M.A., Downey, R.G., Saal, F.E., 1983. Intraclass correlations: there's more there than meets the eye. Psychol. Bull. 93 (3), 586–595.

Li, X., Pan, Y., Fang, Z., Lei, H., Zhang, X., Shi, H., … Rao, H., 2020. Test-retest reliability of brain responses to risk-taking during the balloon analogue risk task. Neuroimage 209, 116495. doi:10.1016/j.neuroimage.2019.116495.

Lipp, I., Murphy, K., Wise, R.G., Caseras, X., 2014. Understanding the contribution of neural and physiological signal variation to the low repeatability of emotion-induced BOLD responses. Neuroimage 86, 335–342. doi:10.1016/j.neuroimage.2013.10.015.

Lois, G., Kirsch, P., Sandner, M., Plichta, M.M., Wessa, M., 2018. Experimental and methodological factors affecting test-retest reliability of amygdala BOLD responses. Psychophysiology 55 (12), 1–15. doi:10.1111/psyp.13220.

Maier, S., Grueschow, M., 2020. Pupil dilation predicts individual success in emotion regulation and dietary self-control. BioRxiv 4534, 1–52. doi:10.1101/2020.11.19.376202.

Marchewka, A., Żurawski, Ł., Jednoróg, K., Grabowska, A., 2014. The nencki affective picture system (NAPS): introduction to a novel, standardized, wide-range, high-quality, realistic picture database. Behav. Res. Methods 46 (2), 596–610 https://doi.org/10.3758/s13428-013-0379-1.

McGraw, K.O., Wong, S.P., 1996. Forming inferences about some intraclass correlation coefficients. Psychol. Methods 1 (1), 30–46.

McRae, K., Gross, J.J., Weber, J., Robertson, E.R., Sokol-Hessner, P., Ray, R.D., … Ochsner, K.N., 2012a. The development of emotion regulation: an fMRI study of cognitive reappraisal in children, adolescents and young adults. Soc. Cogn. Affect. Neurosci. 7 (1), 11–22. doi:10.1093/scan/nsr093.

McRae, K., Jacobs, S.E., Ray, R.D., John, O.P., Gross, J.J., 2012b. Individual differences in reappraisal ability: links to reappraisal frequency, well-being, and cognitive control. J. Res. Personal. 46 (1), 2–7. doi:10.1016/j.jrp.2011.10.003.

McRae, K., Misra, S., Prasad, A.K., Pereira, S.C., Gross, J.J., 2012c. Bottom-up and top-down emotion generation: implications for emotion regulation. Soc. Cogn. Affect. Neurosci. 7 (3), 253–262. doi:10.1093/scan/nsq103.

Merboldt, K.D., Fransson, P., Bruhn, H., Frahm, J., 2001. Functional MRI of the human amygdala? Neuroimage 14 (2), 253–257. doi:10.1006/nimg.2001.0802.

Moeller, S., Yacoub, E., Olman, C.A., Auerbach, E., Strupp, J., Harel, N., Uğurbil, K., 2010. Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain FMRI. Magn. Reson. Med. 63 (5), 1144–1153. doi:10.1002/mrm.22361.

Morawetz, C., Alexandrowicz, R. W., Heekeren, H. R., 2017. Successful emotion regulation is predicted by amygdala activity and aspects of personality: A latent variable approach. Emotion 17 (3), 421–441. doi:10.1037/emo0000215.

Morawetz, C., Bode, S., Baudewig, J., Heekeren, H.R., 2017a. Effective amygdala-prefrontal connectivity predicts individual differences in successful emotion regulation. Soc. Cogn. Affect. Neurosci. 12 (4), 569–585. doi:10.1093/scan/nsw169.

Morawetz, C., Bode, S., Baudewig, J., Jacobs, A.M., Heekeren, H.R., 2016a. Neural representation of emotion regulation goals. Hum. Brain Mapp. 37 (2), 600–620. doi:10.1002/hbm.23053.

Morawetz, C., Bode, S., Baudewig, J., Kirilina, E., Heekeren, H.R., 2016b. Changes in effective connectivity between dorsal and ventral prefrontal regions moderate emotion regulation. Cereb. Cortex 26 (5), 1923–1937. doi:10.1093/cercor/bhv005.

Morawetz, C., Bode, S., Derntl, B., Heekeren, H.R., 2017b. The effect of strategies, goals and stimulus material on the neural mechanisms of emotion regulation: a meta-analysis of fMRI studies. Neurosci. Biobeh. Rev. 72, 111–128. doi:10.1016/j.neubiorev.2016.11.014.

Morawetz, C., Holz, P., Lange, C., Baudewig, J., Weniger, G., Irle, E., Dechent, P., 2008. Improved functional mapping of the human amygdala using a standard functional magnetic resonance imaging sequence with simple modifications. Magn. Reson. Imaging 26 (1), 45–53. doi:10.1016/j.mri.2007.04.014.

Morawetz, C., Kellermann, T., Kogler, L., Radke, S., Blechert, J., Derntl, B., 2016c. Intrinsic functional connectivity underlying successful emotion regulation of angry faces. Soc. Cogn. Affect. Neurosci. 11 (12), 1980–1991. doi:10.1093/scan/nsw107.

Morawetz, C., Riedel, M.C., Salo, T., Berboth, S., Eickhoff, S.B., Laird, A.R., Kohn, N., 2020. Multiple large-scale neural networks underlying emotion regulation. Neurosci. Biobehav. Rev. 116, 382–395. doi:10.1016/j.neubiorev.2020.07.001.

Morris, L.S., Kundu, P., Costi, S., Collins, A., Schneider, M., Verma, G., … Murrough, J.W., 2019. Ultra-high field MRI reveals mood-related circuit disturbances in depression: a comparison between 3-Tesla and 7-Tesla. Transl. Psychiatry (1) 9. doi:10.1038/s41398-019-0425-6.

Moser, E., Stahlberg, F., Ladd, M.E., Trattnig, S., 2012. 7-T MR-from research to clinical applications? NMR Biomed. 25 (5), 695–716. doi:10.1002/nbm.1794.

Neubauer, A.C., Fink, A., 2009. Intelligence and neural efficiency. Neurosci. Biobehav. Rev. 33 (7), 1004–1023. doi:10.1016/j.neubiorev.2009.04.001.

Noble, S., Scheinost, D., Constable, R.T., 2019. A decade of test-retest reliability of functional connectivity: a systematic review and meta-analysis. Neuroimage 203, 116157 https://doi.org/10.1016/j.neuroimage.2019.116157.

Morawetz, C., Berboth, S., Bode, S., 2021. With a little help from my friends: The effect of social proximity on emotion regulation-related brain activity. NeuroImage 230, 117817. doi:10.1016/j.neuroimage.2021.117817.

Noble, S., Scheinost, D., & Constable, R.T. (2020). A guide to the measurement and interpretation of fMRI test-retest reliablity. https://doi.org/https://doi.org/10.31219/osf.io/w3qkf

Ochsner, K.N., Ray, R.D., Cooper, J.C., Robertson, E.R., Chopra, S., Gabrieli, J.D., Gross, J.J., 2004. For better or for worse: neural systems supporting the cognitive down- and up-regulation of negative emotion. Neuroimage 23 (2), 483–499. doi:10.1016/j.neuroimage.2004.06.030.

Ochsner, K.N., Silvers, J., Buhle, J.T., 2012. Functional imaging studies of emotion regulation: a synthetic review and evolving model of the cognitive control of emotion. Ann. N. Y. Acad. Sci. 1251, E1–24. doi:10.1111/j.1749-6632.2012.06751.x.

Otto, B., Misra, S., Prasad, A., McRae, K., 2014. Functional overlap of top-down emotion regulation and generation: an fMRI study identifying common neural substrates between cognitive reappraisal and cognitively generated emotions. Cogn. Affect. Behav. Neurosci. 14 (3), 923–938. doi:10.3758/s13415-013-0240-0.

Paret, C., Hendler, T., 2020. Live from the "regulating brain": harnessing the brain to change emotion. Emotion 20 (1), 126–131. doi:10.1037/emo0000674.

Phan, K.L., Liberzon, I., Welsh, R.C., Britton, J.C., Taylor, S.F., 2003. Habituation of rostral anterior cingulate cortex to repeated emotionally salient pictures. Neuropsychopharmacology 28 (7), 1344–1350. doi:10.1038/sj.npp.1300186.

Picó-Pérez, M., Radua, J., Steward, T., Menchón, J.M., Soriano-Mas, C., 2017. Emotion regulation in mood and anxiety disorders: a meta-analysis of fMRI cognitive reappraisal studies. Prog. Neuropsychopharmacol. Biol. Psychiatry 79, 96–104. doi:10.1016/j.pnpbp.2017.06.001, (May).

Plichta, M.M., Schwarz, A.J., Grimm, O., Morgen, K., Mier, D., Haddad, L., … Meyer-lindenberg, A., 2012. Test – retest reliability of evoked BOLD signals from a cognitive – emotive fMRI test battery. Neuroimage 60 (3), 1746–1758. doi:10.1016/j.neuroimage.2012.01.129.

Powers, J.P., Graner, J.L., Labar, K.S., 2020. Multivariate patterns of posterior cortical activity differentiate forms of emotional distancing. Cereb. Cortex 30 (5), 2766–2776. doi:10.1093/cercor/bhz273.

R CoreTeam. (2016). R: A language and environment for statistical computing. Vienna, Austria.

Ramsey, N.F., Jansma, J.M., Jager, G., Van Raalten, T., Kahn, R.S., 2004. Neurophysiological factors in human information processing capacity. Brain 127 (3), 517–525. doi:10.1093/brain/awh060.

Reinecke, A., Filippini, N., Berna, C., Western, D.G., Hanson, B., Cooper, M.J., … Harmer, C.J., 2015. Effective emotion regulation strategies improve fMRI and ECG markers of psychopathology in panic disorder: implications for psychological treatment action. Transl. Psychiatry 5. doi:10.1038/tp.2015.160, e673.

Revelle, W., 2019. psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern University, Evanston, Illinois.

Robinson, S., Windischberger, C., Rauscher, A., Moser, E., 2004. Optimized 3 T EPI of the amygdalae. Neuroimage 22 (1), 203–210 https://doi.org/10.1016/j.neuroimage.2003.12.048.

Roiser, J.P., Elliott, R., Sahakian, B.J., 2012. Cognitive mechanisms of treatment in depression. Neuropsychopharmacology 37 (1), 117–136. doi:10.1038/npp.2011.183.

Sauder, C.L., Hajcak, G., Angstadt, M., Phan, K.L., 2014. Test-retest reliability of amygdala response to emotional faces. Psychophysiology 50 (11), 1–20. doi:10.1111/psyp.12129.Test-Retest.

Schweizer, S., Grahn, J., Hampshire, A., Mobbs, D., Dalgleish, T., 2013. Training the emotional brain: improving affective control through emotional working memory training. J. Neurosci. 33 (12), 5301–5311. doi:10.1523/JNEUROSCI.2593-12.2013.

Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations : uses in assessing rater reliability. Psychol. Bull. 86 (2), 420–428.

Silvers, J.A., Guassi Moreira, J.F., 2019. Capacity and tendency: a neuroscientific framework for the study of emotion regulation. Neurosci. Lett. 693, 35–39. doi:10.1016/j.neulet.2017.09.017.

Silvers, J.A., Insel, C., Powers, A., Franz, P., Helion, C., Martin, R.E., … Ochsner, K.N., 2017. VlPFC-vmPFC-amygdala interactions underlie age-related differences in cognitive regulation of emotion. Cereb. Cortex 27 (7), 3502–3514. doi:10.1093/cercor/bhw073.

Silvers, J.A., Weber, J., Wager, T.D., Ochsner, K.N., 2015. Bad and worse: neural systems underlying reappraisal of high-and low-intensity negative emotions. Soc. Cogn. Affect. Neurosci. 10 (2), 172–179. doi:10.1093/scan/nsu043.

Sladky, R., Baldinger, P., Kranz, G.S., Tröstl, J., Höflich, A., Lanzenberger, R., … Windischberger, C., 2013. High-resolution functional MRI of the human amygdala at 7 T. Eur. J. Radiol. 82 (5), 728–733. doi:10.1016/j.ejrad.2011.09.025.

Sloan, E., Hall, K., Moulding, R., Bryce, S., Mildred, H., Staiger, P.K., 2017. Emotion regulation as a transdiagnostic treatment construct across anxiety, depression, substance, eating and borderline personality disorders: a systematic review. Clin. Psychol. Rev. 57, 141–7358.

Triantafyllou, C., Hoge, R.D., Krueger, G., Wiggins, C.J., Potthast, A., Wiggins, G.C., Wald, L.L., 2005. Comparison of physiological noise at 1.5 T, 3 T and 7 T and optimization of fMRI acquisition parameters. Neuroimage 26 (1), 243–250 https://doi.org/10.1016/j.neuroimage.2005.01.007.

Uğurbil, K., Adriany, G., Andersen, P., Chen, W., Garwood, M., Gruetter, R., … Zhu, X.H., 2003. Ultrahigh field magnetic resonance imaging and spectroscopy. Magn. Reson. Imaging 21 (10), 1263–1281. doi:10.1016/j.mri.2003.08.027.

Urry, H.L., Van Reekum, C.M., Johnstone, T., Kalin, N.H., Thurow, M.E., Schaefer, H.S., … Alexander, A.L., 2006. Amygdala and ventromedial prefrontal cortex are inversely coupled during regulation of negative affect and predict the diurnal pattern of cortisol secretion among older adults. J. Neurosci. 26 (16), 4415–4425.

Van Der Velde, J., Opmeer, E.M., Liemburg, E.J., Bruggeman, R., Nieboer, R., Wunderink, L., Aleman, A., 2015. Lower prefrontal activation during emotion regulation in subjects at ultrahigh risk for psychosis: an fMRI-study. NPJ Schizophr. 1 (1), 1–7. doi:10.1038/npjschz.2015.26.

Vanderhasselt, M.A., Baeken, C., Van Schuerbeek, P., Luypaert, R., De Raedt, R., 2013. Inter-individual differences in the habitual use of cognitive reappraisal and expressive suppression are associated with variations in prefrontal cognitive control for emotional information: an event related fMRI study. Biol. Psychol. 92 (3), 433–439. doi:10.1016/j.biopsycho.2012.03.005.

Vul, E., Harris, C., Winkielman, P., Pashler, H., 2009. Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. Perspect. Psychol. Sci. 4 (3), 319–324. doi:10.1111/j.1745-6924.2009.01132.x.

Wager, T.D., Davidson, M.L., Hughes, B.L., Lindquist, M.A., Ochsner, K.N., 2008. Prefrontal-subcortical pathways mediating successful emotion regulation. Neuron 59 (6), 1037–1050. doi:10.1016/j.neuron.2008.09.006.

Woo, C.W., Chang, L.J., Lindquist, M.A., Wager, T.D., 2017. Building better biomarkers: brain models in translational neuroimaging. Nat. Neurosci. 20 (3), 365–377. doi:10.1038/nn.4478.

Wright, C.I., Fischer, H., Whalen, P.J., McInerney, S.C., Shin, L.M., Rauch, S.L., 2001. Differential prefrontal cortex and amygdala habituation to repeatedly presented emotional stimuli. Neuroreport 12 (2), 379–383. doi:10.1097/00001756-200102120-00039.

Zilverstand, A., Parvaz, M.A., Goldstein, R.Z., 2017. Neuroimaging cognitive reappraisal in clinical populations to define neural targets for enhancing emotion regulation. A systematic review. Neuroimage 151, 105–116. doi:10.1016/j.neuroimage.2016.06.009.